



117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**JOURNAL COMMENT**  
Unpublished (2001)  
Contact: Singh, J.A.  
Eastern Cereal and Oilseed Research Centre  
Agriculture and Agri-Food Canada  
960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada

	FEATURES	
	SOURCE	
	location/Dualifiers 1..972 /organism="Zea mays" /cultivar="C0328" /db_xref="taxon:4577" /clone="zm04_02c05" /_clone_lib="zm04_AAFRC_ECORC_cold_stressed_maize_seedlings" /_tissue_type="leaf, crown" /_note="Vector: Bluescript SK-/XhoI-ECORI; site_1: Eco RI; site_2: Xho I; Lower temperature 50 C / hour from 22 to 12oc.; bring to 5o in 1 hour from 12oc. leave at 5oc 2 days , photoperiod 16 hours. Light intensity was 125 uE·l <sup>-1</sup> . Library prepared by In vivo mass excision from amplified library."	
BASE COUNT	188 a        326 c        272 g        181 t              5 others	ORIGIN
Query Match	12.0%; Score 120; DB 152; Length 972;	
Best Local Similarity	55.1% ; Pred.No.1.8e-13;	
Matches 283; Conservative 1; Mismatches 216; Indels 14; Gaps 2;		
Oy	412 gtgcggagccgacagggttgatcacaggaacttgtttgtagaccattgcccgaggagtgtgctgcg	471
Db		
Oy	685 GTGTCGTGAAGAAAAGTTGTACTCGGGCTGCACGATGGGTCCACCAGAGTCAAGGCCG	626
Oy	472 gagcgaccttccttcctgcggcgcgcatgtgtccagcccgsaaqticgacagaacgcagcta	531
Db	625 GTTGCGCACGGCCCCAAGGCGCTCGGTAAAMCTSTSGCGGACAACCATGTGGAGATGCCCAATA	566
Oy	532 gaagacattgcogctgtgcagcaagagctgttcataitgactgtccacaacactgtltccccagcgc	591
Db	565 GAAGGCCAACGCCCTGTGTGATGATGACCCAATTGATGGGGCGACACCCTCTCCTGACAGGGCGT	506
Oy	552 ggaccggttcgatgtgtgcatcttgttagaatgtcgaattgtgttcgacagcccaaactgtgcag	651
Db	505 GGAGGCGTCGGGGCGGTGCAGTAACACGACGCACTGATGCCAATGTCAGAGCGCGGAG	446
Oy	652 cgatccctcgcagagagagagatatgtgcgcgcgcgcgaagacccgcgttctgttaacgcgttc	711
Db	445 CGAGGCGCGGAGGCCCTTGACACACTGTCTTGGCGGAGAAGGCCCTTGTCGTTGGGGCCCTTG	386
Oy	712 gctcatctcgcgcgcgaacatttgttgcacagaacagtlgtctctcgcgcttccgcgcgcctg	771

[illegible]

```

VERSION      BE355380.1  GI:9296482
KEYWORDS     EST.
SOURCE       Sorghum.
ORGANISM     Sorghum bicolor
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoideae; Andropogoneae; Sorghum.
              1 (bases 1 to 543)
REFERENCE    Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
AUTHORS      L.H.
TITLE        An EST database from Sorghum: dark-grown seedlings
JOURNAL      Unpublished (2000)
COMMENT      Contact: Cordonnier-Pratt MM
              Department of Botany
              The University of Georgia
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 542 1805
              Email: mmpratt@uga.edu
              Sequences have been trimmed to exclude PolyA, vector and regions
              below phred quality 16. The threshold for highest quality sequence
              is 20.
              Seq primer: JEN REV
              High quality sequence stop: 466
              POLYA-No.

FEATURES
  source      1..543
               /organism="Sorghum bicolor"
               /db_xref="taxon:4558"
               /clone_lib="Dark Grown 1 (DG1)"
               /note="Organ: 5-day-old dark-grown seedlings; Vector:
               Lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was
               made from poly-A RNA in the cloning vector lambda ZAP II.
               Clones to be sequenced were prepared by mass excision."

BASE COUNT   89 a      190 c      182 g      82 t

ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

Query Match      11.7%; Score 116.4; DB 166; Length 543;
Best Local Similarity 55.1%; Pred. No. 8.7e-13;
Matches 300; Conservative 0; Mismatches 226; Indels 18; Gaps 3;

439 ctgtgtggaacacatgcccaggagatgagcgagcggtctcccgcgagcgat 498
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
539 CTGCTCAGCATGGTCCGACGAGTGCAGCCGGTTGGCGACAGCCGCTCGGTAT 480
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499 gtgcagcccgagatgacagcgacgacgacgacgacgacgacgacgacgac 558
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479 TTGCTGCGCGACCACTCGAGAGTGCCCCAGTACAGAGCCCATCTGTCGATGACCA 420
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
559 gtccatgacctgcacacctgctccacagcgcgacgacgacgacgacgacgac 618
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
419 GTTCATGGCGGACGCTTCCTCGACGGGCGTGAGGCGTCGGGGCGATGAGAC 360
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 gaagtcagatgtgcagacccagcctgcagcagcagcagcagcagcagcagcag 678
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 GAGCTCAGCACTAGTCATGTCAGAGCGCGAGAGCGAGCGCCCTCGACGATGTG 300
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
679 ccgagccgaagcccgctgttcgttgacgagcgtcatcctcgccgagcagcgtgc 738
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 CTTCGGGAGAGACCTTTGCTGGGGCCCGGGCGC---CCCAAGAGACCTTGTTGA 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
739 caagcaggtgtcctcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 798
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 GATACACAGCTGGAGGGGCCAGCCAGTCCCGATGGCCCTGCCCAAGATCTCTCTC 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
799 ggtgtgacctgtagagccgacgacgacgacgacgacgacgacgacgacgacgac 858
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 GGCGCGCGCGCTT-----GGCGTAGACCTCGGCGTTGCGAAGAGTTGAGGCC 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
859 gcggtcccgagcggtgtgcacagcagcgagcagcgagcagcgagcagcgagcag 918
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 CGCATCCCGGACGCTGTCAGAGCGCTTGCGCTCTTACGCTCAGCTGTTGTCGAA 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Query Match      10.7%; Score 106.6; DB 166; Length 489;
Best Local Similarity 55.8%; Pred. No. 6.3e-11;
Matches 274; Conservative 0; Mismatches 199; Indels 18; Gaps 3;

492 cggagatgtccacgcccagagtgtagacgacgacgacgacgacgacgacgac 551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 CGGATGATCTGCTGCGCGACCACTCGAGTGCCTCCAGTAAAGGCCCATCTCGTCA 427
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
552 cgaagcgtcatatgctgcacacctgctccacgagcgagcgagcgagcgagcgag 611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 TGACCCAGTTTCATATGCGCGACCGCTCTCTCGACGGCGTGGAGCGCTCGGGG 367
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
612 tctgtgacgctgcatgttgtagacgcccagcagcagcagcagcagcagcagcag 671
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 AGTAGAGAGAGTGCAGTACATGTCAGAGCGCGCGAGAGAGAGCGCGCTCGA 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
672 tgatgtcgcgccgacagcccgctgtgtgacgagcgtcatcctgcgcgcgcgc 731
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 CGATGCTCTTCGCGAGAGACCTTTGCTGGGGCCCGGGC---CGCCCAAGAGACT 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

FEATURES
  source      1..489
               /organism="Sorghum bicolor"
               /db_xref="taxon:4558"
               /clone_lib="Dark Grown 1 (DG1)"
               /note="Organ: 5-day-old dark-grown seedlings; Vector:
               Lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was
               made from poly-A RNA in the cloning vector lambda ZAP II.
               Clones to be sequenced were prepared by mass excision."

BASE COUNT     81 a      168 c      166 g      74 t

ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```







GPO Box 1600, Canberra, ACT, Australia  
Tel: 61 2 6246 5054  
Fax: 61 2 6246 5000  
Email: bryance@pl.csiro.au.

# FEATURES

## source

1. 840  
/organism="Triticum aestivum"  
/cultivar="Wynona"  
/db\_xref="taxon:4565"  
/clone\_lib="P59-1G"  
/clone\_lib="BRX"  
/cell\_type="endosperm"  
BASE COUNT 172 a 266 c 250 g 150 t 2 others  
ORIGIN

Query Match 10.0%; Score 100.2; DB 116; Length 840;  
Best Local Similarity 53.8%; Pred. No. 1e-09;  
Matches 302; Conservative 0; Mismatches 240; Indels 19; Gaps 4;

QY 412 gtccgagacgcaggtgtatagacgactgtgtggaagacatgcccagagatgagcg 471  
DB 655 gtccgagacgcaggtgtatagacgactgtgtggaagacatgcccagagatgagcg 597  
QY 472 ggcggcggtctctctgagcgagcgatgtgcaagcccggaagttcagagcgagcga 531  
DB 536 gtttgcacaccccccacgagcgtccgtaatcgtgcccggggtccactcgtgctcccccagta 537  
QY 532 ggaagacgttcgcgcgtgagcagagcgtgcatggtccatggtccacactgtccacagcgcg 591  
DB 536 gaaaccccccacgtgagacatgac 477  
QY 592 ggaacggtcagatgtgtgcatcgtgtgtagacgtcagatgtgtgtgagcagccagcgtcgacg 651  
DB 476 cctggcgtcgcgcgcgtgagcagac 417  
QY 652 gcatccctcgcagagagagagatgtgagcagcagcagcagcagcagcagcagcagcagcagc 711  
DB 416 cgaagccttgagagccctcagcagatgctggtgagagagagagagagagagagagagagagag 357  
QY 712 gctcatctgcgcgcagacgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 771  
DB 356 gc---gcgcgcagagagagcgtgtgagagagagagagagagagagagagagagagagagagag 300  
QY 772 ggcag 831  
DB 299 ccgagatcgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 252  
QY 832 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 891  
DB 251 ctgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 192  
QY 892 gtgcgtcgtctcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 948  
DB 191 ctgcgtcgtctcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 132  
QY 949 gagcag 969  
DB 131 ccgagagcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 111

## RESULT

8  
BG556798/c 453 bp mRNA EST 10-APR-2001  
LOCUS EML\_38.g03.bl\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
DEFINITION sequence.  
ACCESSION BG556798  
VERSION BG556798.1 GI:13585796  
KEYWORDS EST.  
SOURCE sorghum.  
ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Clade: Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 453)  
Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A., and Pratt, L.H.  
An EST database from Sorghum: developing embryos  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 449  
POLYA-No.

## FEATURES

### source

1. 453  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Embryo 1 (EM1)"  
/note="Organ: Embryos germinated for 24 hr; Vector:  
pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:  
EcoRI; The library was made from PolyA RNA in the cloning  
vector Lambda Zap II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 73 a 158 c 154 g 68 t  
ORIGIN

Query Match 10.0%; Score 99.6; DB 155; Length 453;  
Best Local Similarity 56.4%; Pred. No. 1.4e-09;  
Matches 255; Conservative 0; Mismatches 179; Indels 18; Gaps 3;

QY 521 ggcgcag 580  
DB 452 gtgcgcag 393  
QY 581 tcccaagc 640  
DB 392 tgcagc 333  
QY 641 agcctgc 700  
DB 332 agc 273  
QY 701 ttgagc 760  
DB 272 ttggcgcccccgc 216  
QY 761 ccgc 820  
DB 215 cagccag 164  
QY 821 cag 880  
DB 165 --ggcgtagacccctggtgag 108  
QY 881 aggcgcag 937  
DB 107 agcgccttgccctcctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 48  
QY 938 agcgccttgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 969  
DB 47 agcgtgctcag 16

## RESULT

9  
BG464289/c 549 bp mRNA EST 20-MAR-2001  
LOCUS EML\_71.H11.bl\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
DEFINITION

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
sequence.									
BC6464289									
BC6464289.1	GI:13392642								
EST.									
Sorghum.									
Sorghum bicolor									
Euryotia, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;									
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC									
clade: Penicillidae; Andropogoneae; Sorghum.									
1 (bases 1 to 549)									
Reid, S.P., Cordomier-Pratt, M.-M., Gingle, A. and Pratt, L.H.									
An EST database from Sorghum: developing embryos									
Unpublished (2000)									
Contact: Cordomier-Pratt MM									
Department of Botany									
The University of Georgia									
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA									
Tel: 706 542 1860									
Fax: 706 542 1805									
Email: emprat@uga.edu									
Sequences have been trimmed to exclude polyA, vector and regions									
below phred quality 16. The threshold for highest quality sequence									
is 20.									
Seq primer: JEN REV									
High quality sequence stop: 470									
POLYA-No.									
Location/Qualifiers									
1..549									
/organism="Sorghum bicolor"									
/db_xref="taxon:4558"									
/clone_lib="Embryo 1 (EM1)"									
/note="Organ: Embryos germinated for 24 hr; Vector:									
pBluescript II from lambda Zap II; Site_1: XhoI; Site_2:									
EcoRI; The library was made from polyA RNA in the cloning									
vector lambda Zap II. Clones to be sequenced were									
prepared by mass excision."									
BASE COUNT	92 a	189 c	182 g	85 t	1 others				
ORIGIN									
Query Match	9.9%	Score 99;	DB 154;	Length 549;					
Best Local Similarity	55.5%;	Pred. No. 1.7e-09;							
Matches	261;	Conservative	0;	Mismatches 191;	Indels 18;	Gaps 3;			
513	agttgacgagccgacgctgaagagaccttgcgctgcgacgagctgtccatgctcgtgc								
548	actnagagctgcccagtagagagcccatccctggtcatgacccaagttcagggccga								
573	acaactctgccacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg								
488	ccgctctcttcacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg								
633	cgaagccagacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg								
428	ccatgttcagagccgacgacgacgacgacgacgacgacgacgacgacgacgacg								
693	cgcctctcttcagagccgacgacgacgacgacgacgacgacgacgacgacgacg								
368	cccttctcttcagagccgacgacgacgacgacgacgacgacgacgacgacgacg								
753	cgcgcctcttcagagccgacgacgacgacgacgacgacgacgacgacgacgacg								
311	agc								
813	agagccgc								
253	-----gacgtagacgcttggcgttgcgaagaaggtgacgcccacgacgacg								
873</									

	RESULT	10	
	LOCUS	BF253449/c	
	DEFINITION	BF253449 mRNA EST 23-FEB-2001	
	ACCESSION	HVMEI00001H06f Hordeum vulgare seedling root EST library HVCDNM0007	
	VERSION	(uncloned and unstressed) Hordeum vulgare cDNA clone	
	KEYWORDS	HVMEI00001H06f, mRNA sequence.	
	SOURCE	BF253449 GI:13116514	
	ORGANISM	EST.	
		Barley.	
		Hordeum vulgare	
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae	
		; Triticeae; Hordeum.	
	REFERENCE	1 (bases 1 to 558)	
	AUTHORS	Wing,R., Close,T.J., Kleinjofe,A., Wise,R., Begum,D., Frisch,D., Yu,	
		T., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo	
		,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and	
		Wood,T.	
	TITLE	Development of a genetically and physically anchored EST resource	
	JOURNAL	for barley genomics	
	COMMENT	Unpublished (2000)	
		On Nov 16, 2000 this sequence version replaced gi:11182650.	
		Contact: Wing RA	
		Clemson University Genomics Institute	
		Clemson University	
		100 Jordan Hall, Clemson, SC 29634, USA	
		Tel.: 864 656 7288	
		Fax: 864 656 4293	
		Email: rwing@clemson.edu	
		Seq primer: AATTAACTCCCTCACTAATAAGG	
		High quality sequence stop: 552.	
	FEATURES	Location/Qualifiers	
	Source	1..558	
		/organism="Hordeum vulgare"	
		/cultivar="Morex"	
		/db_xref="taxon:4513"	
		/clone="HVSMEE0001H06f"	
		/clone_lib="Hordeum vulgare seedling root EST library	
		HVCNMA007 (etiolated and unstressed)"	
		/russus_type="Seedling root"	
		/lab_host="TJc121"	
		/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"	
	BASE COUNT	105 a 197 c 179 g 73 t 4 others	
	ORIGIN		
		Query Match 9.6%; Score 95.6; DB 146; Length 558;	
		Best Local Similarity 55.2%; Pred. No. 7.7e-09;	
		Matches 254; Conservative 0; Mismatches 188; Indels 18; Gaps 3;	
Oy	513	agttcgacgagccgcgttaggaaccttcggccctgscgcagagagtgtccatgyctgcc	572
Dd	549	ACTGCNAGGTGGCCCCAGTAGGAAGGCCCAAGCCTGTATCATATGACCACAGTTCAAGCCTTGA	490
Oy	573	acaactcgtcccacgagccgcgagccggtcgatgttgatcatcgtgtagaqtlcgtatggt	632
Dd	489	CGGTTCCTTCATGAGCGCTGTGCGTGCGTGGGCGGCTTGCAGCACAGTCAGAGTACT	430
Oy	633	cgaagcccaagcctgscgagcgatccctctgcaggaagagatgatgtgcgcgcgcagcgc	692
Dd	429	CCATTCGAGCGCGCTTGAAGGAGCGCTTGAAGGCCCTCGACGATGTGCTTGGCGGAGAAGC	370
Oy	693	cgcttctgttagcgcgtcgtcatalctgcgcgcgcacattgtgtgcgcagcacggtgtcct	752
Dd	369	CGTTTTCGTTGGGGCCCCCGCG--CGCCCCAGAAAGAGCTTGATGAGATGACAGAGCTCGG	313
Oy	753	cgagcgcgttcgagccgaccttgaggccagacactgtccaccagagctctctcggltgtgaccttgt	812

Db	312	CGCGGCGGACGCGGAGGCTCCGGATGCGCTGGCCCATGATCTCCTGGGCGCGGCCCTT--	253
Qy	813	agagccgcacgcccgtacatgctggcggtgtctgaagcagatgtatgcgcgagtcctccggagcgt	872
Db	254	-----GGCGTGAACCTCGGCGCTTGCGAAGATGTGACCCCGGCGCTCGCGGAGG	205
Qy	873	ggtccatcagggcgagcgcgtgctgctccctgagcgcgtccgcctgaagtac---cgtagc	929
Db	204	CCTGGAGGAGGAGACTTGCCCTCTTACGTCGACGTGGTGGCGAAGGTATCCACCGACG	145
Qy	930	cgagccagagcctgctgtgtagcagcgcgcgaagccgcag	969
Db	144	CGTAGGACAGCTGGCTCCACCGAGCGCCGACCGGCGNCAG	105
RESULT	11		
LOCUS	AM677520/c		
DEFINITION	DG1.8.E02.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor CDNA, mRNA		
ACCESSION	AM677520		
VERSION	AM677520.1		
KEYWORDS	EST.		
SOURCE	Sorghum		
ORGANISM	Sorghum bicolor		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
AUTHORS	clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 522)		
COMMENT	Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.		
	An EST database from Sorghum: dark-grown seedlings unpublished (2000)		
	Contact: Cordonier-Pratt MM		

Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@atcuga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequences is 20.  
Seq primer: JEN REV  
High quality sequence stop: 465  
POLYA-No.

FEATURES	SOURCE	Location/Qualifiers
	1.	522
		Location/Qualifiers
		1. 522
		/organism="Sorghum bicolor"
		/db_xref="taxon:4558"
		/clone_1ib="Dark grown 1 (DG1)"
		/note="Organ: 5-day-old dark-grown seedlings; Vector: lambda zap; Site.1: XhoI; Site.2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	86 a	181 c
ORIGIN		1/2 g 83 t

	Query Match	Similarity	9.3%	Score 93.2	DB 119	Length 522
	Best Local	Similarity	56.2%	Pred. No. 2.2e-08		
	Matches	245	Conservative	0	Mismatches 173	Indels 18; Gaps 3
OY	547	ggcgcgcgagcgtctcatatgctctgcacacccctcgtcccaagcgcgagccggtcgtatgt	606			
Db	522	gttcgatgacccagtttcattatggccgcacccctcttccttcgacggcggtgagagcgttcggggcg	463			
OY	607	gtgcattctgtagtaagtcgatgtgtrtcgacgcccagctcgtgcagagatattccctcgcagga	666			
Db	462	atggcagtagtagacagacttcacagcttagtccattgttcgaagcgccggagagccggcgaagcc	403			
OY	667	ggcgaatgtagtgcgcgcgcgcagaccccgcgtgttcgttgagcgcgtgcgtcatcttcgcgcgc	726			

Db	402	CTGACGATGTGGCTTCCGGGAGAGACCCCTTTCGTTGGGGGCGCGGGC---CGCCCCAGAA	346
QY	727	gaacttgatcgccagcaacgylgtcctcgcgccgctccgcgcgcctcggccagcaacatcgcc	786
Db	345	GAGCTGGTGGAGATGAGACGAGCTGGACCGCGCCAGCCACAGTGTCCGGATGGCGCTGCC	286
QY	787	caccaagccctccgylgtctgcacctgtatagaacgcgcacagccagccgatactgcagcgylgtcga	846
Db	285	CATGATCTCTCTCGGGGCGCCGTT-----GGCGTAGACCTCTGGCGTTGTGGA	238
QY	847	gcagttgatgcgcgyltccggcggtgtgtcatatagcgcagcgcgylgtcgtcctcgac	906
Db	237	GAGGTTGACGCCCGCATCCCGGACAGCGCTGACGACGAGCGCTTGCTTCACGTCCAG	178
QY	907	gcgtccgcttgaaatctaac---cgttcgcagacagagcctctgtgtgacagcgcgagac	963
Db	177	CTGTTGGCCGAAGGTCACCCACGAGCCGTACGAGAGCTGCGTCCACGCGAGCCCCGACCG	118
QY	964	cccgagccgcagcagctgc	979
Db	117	GCCCAGGTTCTTTGTAC	102

RESULT	12
BE775466/C	
LOCUS	
DEFINITION	BE775466 629 bp mRNA
ACCESSION	MY-01-C-01 PinfastansMW Phytophthora infestans cDNA, mRNA seq
VERSION	BE775466
KEYWORDS	BE775466.1 GI:10229009
SOURCE	EST
ORGANISM	potato late blight agent. Phytophthora infestans

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1 (bases 1 to 629)	Kamoun, S., Habber, P.T., Sobral, B.W.S., Nuss, D. and Govers, F.	Initial assessment of gene diversity for the oomycete pathogen <i>Phytophthora infestans</i> based on expressed sequences	Fungal Genet. Biol. 28 (2), 94-106 (1999)	20056376	Contact: Govers F
	Euariyola, Stramenopiles; Oomycetes; Pycnatales; Pycnataceae; Phytophthora.				

Laboratory of Phytopathology  
Wageningen University  
Bilthoven 9, P.O. Box 8025, 6700 EE, Wageningen, The Netherlands  
Tel.: 31 317 483 138  
Fax: 31 317 483 412  
Email: [Francine.Govers@medew.fyvo.wau.nl](mailto:Francine.Govers@medew.fyvo.wau.nl)  
Location/Qualifiers  
1. 629

/organism="Phycoplocus infestans"  
 /strain="DDR/602, AI mating type"  
 /db\_xref="taxon:4787"  
 /clone\_lib="PinfestansMY"  
 /dev\_stage="4-week old vegetative, non-sporulating  
 mycelium in synthetic medium"  
 /lab\_host="E. coli, strain DH5-alpha"  
 /note="Vector: pSPORT1; Site\_1: SalI; site\_2: NotI; T  
 RNA was isolated from mycelium of P. infestans DDR/606-  
 cultured for 4 weeks in synthetic medium. EST clones  
 named by their position in the microtiter plate, pre-  
 by the prefix MY (for mycelial) and the successive n  
 of the microtiter plate (e.g. MY-06-A-04)."  
 144 a 185 c 158 g 126 t 16 others

	Query Match	9.3%	Score 93	DB 140	Length 629
	Best Local Similarity	47.5%	Pred. No. 2.4e-08		
	Matches 252	Conservative 0	Mismatches 279	Indels 0	Gaps
Oy	425 aggtgtatcaagcactcgtgtggaacatcagcccgagatggtggcgcgcggtcttc				484
Db	609 atggagatcctgtggtgttaacacggagcggtgcagacccgacccgctgcgatnwcg				550

[illegible]

RESULT	13				
LOCUS	BE494103/c				
DEFINITION	BE494103	596 bp	mRNA	EST	02-AUG-2000
ACCESSION	WHE1277.H02.P03S	Secale cereale	anther cDNA	library	Secale cereale
VERSION	BE494103				
KEYWORDS	BE494103.1	GI:9660696			
SOURCE	rye.				
ORGANISM	Secale cereale				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Secale.				
AUTHORS	1 (bases 1 to 596)				
TITLE	Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Ross,K., Seaton,C.B. and Tong,J.C.				
JOURNAL	The structure and function of the expressed portion of the wheat genomes - Anther cDNA library from rye				
COMMENT	Unpublished (2000)				
CONTACT	Contact: Olin Anderson				
WEST AREA	US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center				
TEL	800 Buchanan Street, Albany, CA 94710, USA				
FAX	Tel: 5105595773				
EMAIL	Fax: 5105595818				
SEQ PRIMER	Email: coanders@wpr.usda.gov				
QUALITY	Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20				
LOCATION	Seq primer: Stratiogene SK primer.				
QUALIFIERS	Location/Qualifiers				
SOURCE	1..596				

```

/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WH8127_H02_P03"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"

```

/lab host="E. coli SOLR"  
 /note="Vector: Lambda uni-ZAP XR, excised phagemid;  
 Site\_1: EcoRI; Site\_2: XhoI; Plants were grown in the  
 greenhouse. Anthers were harvested and pooled from early  
 meiosis to late meiosis. The tissue, total RNA, and  
 poly(A) RNA were prepared (Butler, Ross and Gustafson) at  
 University of Missouri, Columbia. A cDNA library was  
 made, and the cDNA clones were in vivo excised to give  
 phlucscript phagemids in the TJ Clouse lab (Choi, Clouse,  
 Fenton) at the University of California, Riverside.  
 Plasmid DNA preparations and DNA sequencing were performed  
 in the OD Anderson lab (all other authors)."  
 BASE COUNT 103 a 218 c 181 g 94 t  
 ORIGIN

Query Match	8.3%	Score 83	DB 136	Length 596
Best Local Similarity	53.4%	Pred. NO	1.9e-06	
Matches 229	Conservative	0	Mismatches 185	Indels 15
				Gaps 2

[illegible]

RESULT 14  
AM447980/c  
LOCUS AM447980 675 bp mRNA  
DEFINITION BR\_1163 BRY Triticum aestivum cDNA clone P25-1D, mRNA sequence.  
ACCESSION AM447980  
VERSION  
KEYWORDS AM447980.1 GI:12018515  
SOURCE EST.  
ORGANISM bread wheat.  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae  
; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 675)  
AUTHORS Clarke,B.C., Hobbs,M. and Appels,R.  
TITLE Genes active in developing wheat endosperm  
JOURNAL Unpublished (2000)  
COMMENT Contact: Bryan Clarke  
Division of Plant Industry  
C.S.I.R.O.  
GPO Box 1600, Canberra, ACT, Australia



W

This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 22, 2001, 09:22:44 ; Search time 22.7 Seconds

(without alignments)  
889.330 Million cell updates/sec

Title: US-09-463-705A-2

Perfect score: 1728  
Sequence: 1 MTDDAATHVALGSRALITSR.....LDELFPVAVSGAEPAMWLQ 333Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

A.Geneseq\_0601.\*  
1: /SIDSB/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	100.0	333	AAW19736	Sugar biosynthesis
2	1728	100.0	333	AAW99386	S.erythraea dtdp-4
3	1056	61.1	347	AAB61981	S. avermitilis ORF
4	418	24.2	20	AAV52837	Escherichia coli p
5	380.5	22.0	328	AAAG39049	Arabidopsis thaliana
6	370	21.4	328	AAAG35919	Arabidopsis thaliana
7	356	20.6	329	AAW23388	Zea mays protein f
8	353.5	20.5	412	AAAG41675	Shaker-like potass
9	353.5	20.5	419	AAAG41674	Arabidopsis thaliana
10	347.5	20.1	357	AAAG41676	Arabidopsis thaliana
11	337	19.5	312	AAB79618	Cornebacterium gl

12	337	19.5	312	22	AAB80067
13	331	19.2	329	18	AAW23389
14	329	19.0	344	19	AAW69711
15	321	18.6	344	21	AAB08255
16	317.5	18.4	340	21	AAAG38206
17	317.5	18.4	349	21	AAAG38205
18	317.5	18.4	351	21	AAAG38204
19	301	17.4	337	21	AAAG07306
20	301	17.4	346	21	AAAG07305
21	301	17.4	349	21	AAAG07304
22	300.5	17.4	345	21	AAAG31466
23	299	17.3	266	21	AAAG39050
24	285.5	16.5	319	21	AAAG31467
25	284.5	16.5	266	21	AAAG35920
26	279.5	16.2	327	21	AAAG40256
27	279.5	16.2	318	21	AAAG40255
28	271.5	15.7	281	21	AAAG54764
29	271	15.7	301	21	AAAG40257
30	269	15.6	342	21	AAAG29729
31	265	15.3	315	22	AAW76845
32	261.5	15.1	290	22	AAW76844
33	258.5	15.0	316	21	AAAG29730
34	252	14.6	287	21	AAAG30281
35	252	14.6	296	21	AAAG30280
36	245.5	14.2	316	21	AAAG35251
37	245.5	14.2	371	21	AAAG35250
38	243.5	14.2	388	21	AAAG35249
39	243.5	14.1	270	21	AAAG30282
40	242.5	14.0	375	21	AAAG52050
41	242.5	14.0	377	21	AAAG52049
42	242.5	14.0	386	21	AAAG52048
43	237	13.7	284	21	AAAG31468
44	234	13.5	316	21	AAAG39500
45	234	13.5	324	21	AAAG39499

## ALIGNMENTS

RESULT 1  
ID AAW19736 standard; Protein; 333 AA.  
XX  
AC AAW19736;  
XX  
DT 18-SEP-1997 (first entry)  
XX  
DE Sugar biosynthesis enzyme EryBII.  
XX  
KW Polyketide; glycosylation; eryBII; L-mycarose; antimicrobial;  
KW antibiotic; antifungal; fungicide; anticancer; cytostatic;  
KW anthelmintic.  
XX  
OS Saccharopolyspora erythraea.  
XX  
PN WO9723630-A2.  
XX  
PD 03-JUL-1997.  
XX  
PF 23-DEC-1996; 96WO-US20238.  
XX  
PR 21-DEC-1995; 95US-0576626.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Donadio S, Katz L, Staver MJ, Summers RG;  
XX  
DR WPI. 1997-351066/32.  
XX  
PT N-PSDB; AAT72684.  
XX  
PT New genes involved in sugar biosynthesis and attachment - used to  
glycosylation generate polyketide antimicrobials etc. with altered pattern of

XX Disclosure; Fig 4A; 85bp; English.  
 PS EryBII (AAW19736) is an enzyme involved in the biosynthesis of the  
 CC sugar L-mycarose. It is one of 10 enzymes (see also AAW19734-35 and  
 CC AAW19737-43) predicted to be involved in D-desosamine or L-mycarose  
 CC biosynthesis and attachment that have been identified from gene  
 CC clusters (AAV72684 and AAV72685) of *Saccharopolyspora erythraea*. Novel  
 CC glycosylation-modified polyketides are produced by selectively  
 CC altering, inactivating or augmenting the *eryB* and/or *eryC* genes  
 CC encoding these sugar biosynthesis enzymes and introducing them into  
 CC polyketide-producing microorganisms.  
 CC  
 XX Sequence 333 AA:  
 SQ

Query Match 100.0%; Score 1728; DB 18; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-165;  
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTDAATHVRLGRSALLTSRLMGTVNFSGRVEDDRLMDHARDGINCPLTDADMYGW 60  
 DB 1 mtdaathvrlgrsalltsrlmgtvnfsgvrveddrlmdhadrtdgincldtdadmygw 60  
 QY 61 RLTKGHEELVGRMLAOGGGRREDTVLATKVGEMSERVNDGSLAHHTTASCEGSLRRL 120  
 DB 61 rltkghellevgrmlaogggredvtlatkvgemservndsgslahhttascegsrlrl 120  
 QY 121 GVPHIDVYOMHHIDRSAPMGEVWQAMDLSVASKGVSVSSNFGWHITAAOENAAARHS 180  
 DB 121 gvphidvygmhhidrsapmgevwqamdsivasgkvsyvsngntagwhiaaqaenaarhs 180  
 QY 181 LGMVSHOCLYNLAVRHALEVLPAQAQYGLGVFAMSPDHGGLSGALEKTLAAGTAVKSAQ 240  
 DB 181 lgmshqclynlavrhaelevlpaagqyglgvfawspdhgllsgalektaagtavksaq 240  
 QY 241 GRAOVLPLSRPALEAVEKRCNLGEDPAEYGLAWLSRPGIAGAVIGPRTPEQDSALK 300  
 DB 241 graqvllpslrpaleavekrcnlgedpaevglawlsrpgiagavigprtpeqldsalk 300  
 QY 301 ASAMTIDEOALSELDEFPAVASGGAPEAMIQ 333  
 DB 301 asamtld eaglseldelfpavasggaapeawiq 333

RESULT 2  
 AAW93386  
 ID AAW93386 standard; Protein; 333 AA.  
 AC AAW93386;  
 XX  
 DT 08-JUN-1999 (first entry)  
 XX  
 DE S. erythraea dTDP-4-keto-L-6-deoxyhexose-2,3-reductase.  
 XX  
 KW Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;  
 KW secondary metabolite; eryBII; eryCII; hybridisation; probe;  
 KW glycosylation; macrolactone; oleandomycin.  
 XX  
 OS Saccharopolyspora erythraea.  
 XX  
 PN WO9905283-A2.  
 XX  
 PD 04-FEB-1999.  
 XX  
 PF 21-JUL-1998; 98WO-FR01593.  
 XX  
 PR 12-JUN-1998; 98FR-0007411.  
 PR 25-JUL-1997; 97FR-0009458.  
 XX  
 PA (HMRI ) HOECHST MARION ROUSSEL.  
 XX  
 PI Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC;

PI Salah-Bey K, Fromentin C, Mendez C, Salas JA;  
 XX  
 DR WPI; 1999-142938/12.  
 DR N-PSDB; AAX25772.  
 XX  
 PT New nucleic acid sequences encoding enzymes involved in macrolide  
 PT biosynthesis - useful for producing hybrid secondary metabolites,  
 PT particularly erythromycin analogues  
 PS  
 XX Claim 6; Fig 2; 221bp; French.  
 CC  
 CC This sequence represents the dTDP-4-keto-L-6-deoxyhexose-2,3-reductase  
 CC enzyme encoded by the *eryBII* gene from the *eryE-eryAIII* gene cluster from  
 CC the Gram-positive bacterium *Saccharopolyspora erythraea*. This gene  
 CC cluster encodes enzymes involved in the production of the macrolide  
 CC antibiotic erythromycin as a secondary metabolite. The erythromycin gene  
 CC cluster spans approximately 53 kb and contains at least 20 open reading  
 CC frames (ORF). The genes are used to produce hybrid secondary metabolites  
 CC in *S. erythraea*, i.e. erythromycin analogues which may have improved  
 CC properties or as hybridisation probes for isolating homologous genes  
 CC involved in glycosylation of macrolactones in macrolide-producing  
 CC strains (specifically oleandomycin-producing strains of *Streptomyces*  
 CC antibioticus).  
 CC  
 XX Sequence 333 AA:  
 SQ

Query Match 100.0%; Score 1728; DB 20; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-165;  
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTDAATHVRLGRSALLTSRLMGTVNFSGRVEDDRLMDHARDGINCPLTDADMYGW 60  
 DB 1 mtdaathvrlgrsalltsrlmgtvnfsgvrveddrlmdhadrtdgincldtdadmygw 60  
 QY 61 RLTKGHEELVGRMLAOGGGRREDTVLATKVGEMSERVNDGSLAHHTTASCEGSLRRL 120  
 DB 61 rltkghellevgrmlaogggredvtlatkvgemservndsgslahhttascegsrlrl 120  
 QY 121 GVPHIDVYOMHHIDRSAPMGEVWQAMDLSVASKGVSVSSNFGWHITAAOENAAARHS 180  
 DB 121 gvphidvygmhhidrsapmgevwqamdsivasgkvsyvsngntagwhiaaqaenaarhs 180  
 QY 181 LGMVSHOCLYNLAVRHALEVLPAQAQYGLGVFAMSPDHGGLSGALEKTLAAGTAVKSAQ 240  
 DB 181 lgmshqclynlavrhaelevlpaagqyglgvfawspdhgllsgalektaagtavksaq 240  
 QY 241 GRAOVLPLSRPALEAVEKRCNLGEDPAEYGLAWLSRPGIAGAVIGPRTPEQDSALK 300  
 DB 241 graqvllpslrpaleavekrcnlgedpaevglawlsrpgiagavigprtpeqldsalk 300  
 QY 301 ASAMTIDEOALSELDEFPAVASGGAPEAMIQ 333  
 DB 301 asamtld eaglseldelfpavasggaapeawiq 333

RESULT 3  
 AAB61981  
 ID AAB61981 standard; Protein; 347 AA.  
 AC AAB61981;  
 XX  
 DT 14-MAY-2001 (first entry)  
 XX  
 DE S. avermitilis ORF8 (AvrI) polypeptide.  
 XX  
 KW Glycosylation; avermectin; aglycone; biocatalysis; antibiotic;  
 KW oleandomycin; ORF8; AvrI.  
 XX  
 OS Streptomyces avermitilis.  
 XX  
 PN WO200109155-A1.  
 XX

PD 08-FEB-2001.  
 XX 26-JUL-2000; 2000WO-US20331.  
 PF 30-JUL-1999; 99US-0146699.  
 PR (MERI ) MERCK & CO INC.  
 XX MacNeil DJ, Occhi J, Gewain KM;  
 PI MPI; 2001-182929/18.  
 DR N-PSDB; AAC85191.  
 XX Novel nucleic acid fragments of Streptomyces avermitilis genome useful  
 PT for enzymatic, biochemical, biosynthetic and diagnostic purposes -  
 PS Disclosure; Page 22; 63pp; English.  
 XX The invention relates to a 10 kb genomic DNA isolated from *S. avermitilis*  
 CC that contains genes encoding proteins for glycosylation of avermectin  
 CC aglycones. The polypeptides can be expressed by standard recombinant  
 CC methodology. The gene and the encoded polypeptides together with other  
 CC enzymatically active polypeptides, are useful to perform combinatorial  
 CC biocatalysis in vitro and in vivo in a host cell. They are useful for  
 CC performing biotransformations on macrocyclic compounds including avermectin  
 CC or other macrocyclic aglycones. The sequences are also useful in vivo in a  
 CC bacterial host, in vitro in combination with an actinomycete  
 CC fermentation, and in vitro in combination with an enzymatically active  
 CC polypeptides that are not from the avermectin biosynthetic pathway to  
 CC effect the synthesis of a pharmaceutically active compound, e.g. an  
 CC antibiotic. Sequences AAB61973-981 represent polypeptides in the  
 CC avermectin biosynthetic pathway, involved in the synthesis and/or  
 CC addition of oleandrose to avermectin aglycones. The present sequence is  
 CC the *S. avermitilis* ORF8 (AVR1) polypeptide having the activity of  
 CC oleandrose synthesis.  
 XX Sequence 347 AA;  
 SQ  
 Query Match 61.1%; Score 1056; DB 22; Length 347;  
 Best Local Similarity 61.5%; Pred. No. 1.3e-97;  
 Matches 203; Conservative 35; Mismatches 92; Indels 0; Gaps 0;  
 OY 3 TDAATHVRLGNSALITSRMLCTVNFSGRVEDDDALRLMDHARDGINCILDTADMYGWL 62  
 DB 17 ssaphtlhtlgrtlrlslslalgtvnlgrveepearlmdhlaagilfdantgywrv 76  
 OY 63 YKGHREELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHITLASCESLRLGCV 122  
 DB 77 hkyteevlgrvtdrparrevgvlatkvqdpmsgpndhlsytnlvaacdaslrrlt 136  
 OY 123 DHIDVYOMHHDIDRSAPWDEWQAMDSLVASGKVSYSVSSNFAQWGHIAAQAENARRHSLG 182  
 DB 137 dwidlyghldhrragwdevwgamliltgkvrvygsnfagwdaaqaarrnaly 196  
 OY 183 MWSHOCYLNLAHRAHELEVLPAQAAYGCVFAPWSPLHGGLSGALAEKLAAGYAVKSAQR 242  
 DB 197 laseqcyvnlvtrhaelelvipaasaygyvlywspbhgl19gvlrktrentavksagqr 256  
 OY 243 AQVILPSLRPAIEAEKRCRNLCGDPAEYGLAWLSRPIACVAGVPRPELDSATKMS 302  
 DB 257 avealehhtltaayedvcadngldpanvgnawlsrpsvgtvlgprteghvdaalhal 316  
 OY 303 AMTLDEQALSELDEIFPVAASGAPEAWL 332  
 DB 317 rtlpelpvlarleellfpvgrgsapdaw1 346  
 RESULT 4  
 AAY52837  
 ID AAY52837 standard; Protein; 348 AA.  
 XX AAY52837;  
 AC

XX 26-JAN-2000 (first entry)  
 DT Escherichia coli protein sequence SEQ ID NO:4.  
 DE Isoprenoid: microorganism; detection; antibacterial; herbicide;  
 XX heart disease; osteoporosis; haemostasis; cancer; immunopotentialion;  
 KW health food; antifouling coating; farnesyl pyrolinic acid; pyruvic acid;  
 KW 1-deoxy-D-xylulose-5-phosphate; glyceraldehyde-3-phosphate;  
 KW 2-C-methyl-D-erythreitol-4-phosphate.  
 XX Escherichia coli.  
 XX W09953071-A1.  
 PN 21-OCT-1999.  
 XX 14-APR-1999; 99WO-JP01987.  
 PF 14-APR-1998; 98JP-0103101.  
 PR 05-AUG-1998; 98JP-0221910.  
 PR 15-FEB-1999; 99JP-0035739.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA Miyake K, Hashimoto S, Motoyama H, Ozaki A, Seto H, Kuzuyama T;  
 PI Takahashi S;  
 PI WPI; 1999-620434/53.  
 DR N-PSDB; AA233162.  
 XX Preparation of recombinant isoprenoid compounds useful for treatment of  
 PT heart diseases, osteoporosis and hemostasis, preventing cancer and  
 PT immunopotentialion  
 XX Claim 1; Page 70-72; 145pp; Japanese.  
 PS The present invention describes the preparation of an isoprenoid  
 CC compound comprising using at least 1 DNA e.g. encoding proteins which  
 CC elevate the efficiency of the synthesis or DNA encoding a farnesyl  
 CC pyrolinic producing enzyme. The method of preparation of an isoprenoid  
 CC compound comprises using at least 1 DNA, a vector, cloned cells, their  
 CC derived recombinant DNAs or transformed products in a culture system and  
 CC extracting the isoprenoid accumulated in the medium. The DNA encodes at  
 CC least 1 of the following: (a) a compound for activating or catalysing  
 CC the production of 1-deoxy-D-xylulose-5-phosphate from pyruvic acid and  
 CC glyceraldehyde-3-phosphate; (b) an enzyme producing farnesyl pyrolinic  
 CC acid; (c) a protein which elevates the efficiency of synthesis of  
 CC isoprenoid compounds and comprises a 3 or 4 amino acid sequence  
 CC optionally with 1 or more of the amino acids being deleted or  
 CC substituted or an additional amino acid being inserted; (d) a protein  
 CC which activates or catalyses the production of 2-C-methyl-D-erythreitol-  
 CC 4-phosphate from 1-deoxy-D-xylulose-5-phosphate; or (e) a protein which  
 CC activates a target compound or reaction and is a string end or hybrid of  
 CC the DNA encoded in (a)-(e). Isoprenoid compounds are useful in drugs  
 CC (e.g. for the treatment of heart diseases, osteoporosis and hemostasis,  
 CC antifouling coatings, health foods and  
 CC on the non-mevalonate pathway and can be used as antibacterials and  
 CC herbicides. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX Sequence 348 AA;  
 SQ  
 Query Match 24.2%; Score 418; DB 20; Length 348;  
 Best Local Similarity 34.6%; Pred. No. 1.4e-33;  
 Matches 112; Conservative 52; Mismatches 124; Indels 36; Gaps 7;  
 OY 11 LGRSALITSRMLGTVNF-----SGRVEDDDALRLMDHARDGINCILDTADMYGWL 62  
 DB 30 lgtklrlyrslcgmctfsgpdrghnawtlpessrplkrlagglntfdansys--- 86



PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147182.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149384.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149920.
PR	23-AUG-1999;	99US-0149930.
PR	23-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157119.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158339.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161361.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match		
Best Local Similarity 30.6%; Pred. No. 7.3e-30;		
Matches 102; Conservative 64; Mismatches 144; Indels 23; Gaps		
QY	11 LGRSALLSTRMLGT-VNFSGRVDDDALRLMDIARRGICLDTADMYGKLYKHTFE	69
Db	6 LKSGIKVSTLSIGAWLFGNGLDVKAAKSLIGCCRHGYVFFDAEYVA---NGRAEE	61
QY	70 LVGRMLAOGGGRREDYVATKV--GGEHMERVNDSGTSARHIIASCESGRLGVHDIDV	127
Db	62 IMGGAITELGWRISDVLSTKTLFGG---PGPNDKGJSLRKHIVGKTASIKRLIMDYVDV	118
QY	128 YQMHIDRSAPWDEWQAMDSLVASGVSYGVSSNFGMHIIAAQENARHRSIGVSHQ	187
Db	119 LYCHRPASPRIEETVRAMYUPLDKIGVAFYGTSESAQGLTEAWGADRIDLVPIVEQ	178
QY	188 CLYVLAVRHH-AELEVLRLAAQAVYGVFAMSPFHGGLISGALKEKIAAGT---AVKSQGR	242
Db	179 PEYMMFARHKVETELFPLRYTHVIGLITCWSPLASGVLTGYKXKAIPSDSFALENYKNI	238
QY	243 A-QVLLPSLRPAIEAYEKFCRNLEDPAEYGLAWLVRPGIAGAVIGPRPEQDLSALKA	301
Db	239 ANRSLVDVLRKVSGLPIADELBYVTLAQIALAWCAENPVSSVLTGATESQIGENMKA	298
QY	302 SAMTLDEQALSE---LDEIFPVAVSGGAAPPAW	331
Db	299 ---VDVIFPLTPIDVLDKIEGLQSKPKRPRESY	327

PR	04-MAY-1999;	9905-0132484;	PR	22-JUL-1999;	9905-0145089;
PR	05-MAY-1999;	9905-0132485;	PR	22-JUL-1999;	9905-0145087;
PR	06-MAY-1999;	9905-0132486;	PR	22-JUL-1999;	9905-0145089;
PR	07-MAY-1999;	9905-0132487;	PR	22-JUL-1999;	9905-0145192;
PR	11-MAY-1999;	9905-0134256;	PR	23-JUL-1999;	9905-0145145;
PR	14-MAY-1999;	9905-0134218;	PR	23-JUL-1999;	9905-0145218;
PR	14-MAY-1999;	9905-0134219;	PR	26-JUL-1999;	9905-0145224;
PR	14-MAY-1999;	9905-0134221;	PR	27-JUL-1999;	9905-0145276;
PR	18-MAY-1999;	9905-0134370;	PR	27-JUL-1999;	9905-0145918;
PR	19-MAY-1999;	9905-0134768;	PR	27-JUL-1999;	9905-0145919;
PR	20-MAY-1999;	9905-0134941;	PR	28-JUL-1999;	9905-0145951;
PR	21-MAY-1999;	9905-0135124;	PR	02-AUG-1999;	9905-0146386;
PR	24-MAY-1999;	9905-0135353;	PR	02-AUG-1999;	9905-0146388;
PR	25-MAY-1999;	9905-0136029;	PR	02-AUG-1999;	9905-0146389;
PR	27-MAY-1999;	9905-0136392;	PR	03-AUG-1999;	9905-0147038;
PR	28-MAY-1999;	9905-0136782;	PR	04-AUG-1999;	9905-0147204;
PR	01-JUN-1999;	9905-0137222;	PR	04-AUG-1999;	9905-0147302;
PR	03-JUN-1999;	9905-0137528;	PR	05-AUG-1999;	9905-0147192;
PR	04-JUN-1999;	9905-0137502;	PR	05-AUG-1999;	9905-0147260;
PR	07-JUN-1999;	9905-0137724;	PR	06-AUG-1999;	9905-0147303;
PR	08-JUN-1999;	9905-0138099;	PR	09-AUG-1999;	9905-0147493;
PR	10-JUN-1999;	9905-0138540;	PR	09-AUG-1999;	9905-0147935;
PR	10-JUN-1999;	9905-0138847;	PR	10-AUG-1999;	9905-0148171;
PR	14-JUN-1999;	9905-0139119;	PR	11-AUG-1999;	9905-0148319;
PR	16-JUN-1999;	9905-0139452;	PR	12-AUG-1999;	9905-0148341;
PR	17-JUN-1999;	9905-0139493;	PR	13-AUG-1999;	9905-0148565;
PR	18-JUN-1999;	9905-0139454;	PR	13-AUG-1999;	9905-0148684;
PR	18-JUN-1999;	9905-0139455;	PR	16-AUG-1999;	9905-0149368;
PR	18-JUN-1999;	9905-0139456;	PR	17-AUG-1999;	9905-0149175;
PR	18-JUN-1999;	9905-0139457;	PR	18-AUG-1999;	9905-0149426;
PR	18-JUN-1999;	9905-0139458;	PR	20-AUG-1999;	9905-0149723;
PR	18-JUN-1999;	9905-0139459;	PR	20-AUG-1999;	9905-0149929;
PR	18-JUN-1999;	9905-0139460;	PR	23-AUG-1999;	9905-0149902;
PR	18-JUN-1999;	9905-0139461;	PR	23-AUG-1999;	9905-0150930;
PR	18-JUN-1999;	9905-0139462;	PR	23-AUG-1999;	9905-0150566;
PR	18-JUN-1999;	9905-0139463;	PR	26-AUG-1999;	9905-0150884;
PR	18-JUN-1999;	9905-0139750;	PR	27-AUG-1999;	9905-0151086;
PR	18-JUN-1999;	9905-0139763;	PR	27-AUG-1999;	9905-0151066;
PR	21-JUN-1999;	9905-0139817;	PR	27-AUG-1999;	9905-0151080;
PR	22-JUN-1999;	9905-0139899;	PR	30-AUG-1999;	9905-0151303;
PR	23-JUN-1999;	9905-0140355;	PR	31-AUG-1999;	9905-0151438;
PR	23-JUN-1999;	9905-0140354;	PR	01-SEP-1999;	9905-0151930;
PR	24-JUN-1999;	9905-0140693;	PR	01-SEP-1999;	9905-0152363;
PR	28-JUN-1999;	9905-0140823;	PR	07-SEP-1999;	9905-0153070;
PR	29-JUN-1999;	9905-0140991;	PR	10-SEP-1999;	9905-0153078;
PR	30-JUN-1999;	9905-0141287;	PR	13-SEP-1999;	9905-0154018;
PR	01-JUL-1999;	9905-0141845;	PR	15-SEP-1999;	9905-0154075;
PR	01-JUL-1999;	9905-0142155;	PR	16-SEP-1999;	9905-0154039;
PR	02-JUL-1999;	9905-0142055;	PR	20-SEP-1999;	9905-0154779;
PR	06-JUL-1999;	9905-0142390;	PR	22-SEP-1999;	9905-0155139;
PR	08-JUL-1999;	9905-0142803;	PR	23-SEP-1999;	9905-0155486;
PR	08-JUL-1999;	9905-0142920;	PR	24-SEP-1999;	9905-0155659;
PR	09-JUL-1999;	9905-0142927;			



KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134570.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147392.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148345.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.



U

[illegible]

Db	348	ltpvetaigfvrdrpfvstlligatsvqxklkedaidalimterpfsqewmadidavf	403
RESULT	9		
ID	ACG41674		
XX	ACG41674	standard; Protein; 419 AA.	
AC	ACG41674;		
XX			
DT	18-OCT-2000	(first entry)	
XX			
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 51881.	
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136352.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	17-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139452.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	

PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160880.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161356.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 20.5%; Score 353.5; DB 21; Length 419;  
Best Local Similarity 30.9%; Pred. No. 5, 3e-27;  
Matches 110; Conservative 63; Mismatches 140; Indels 43; Gaps 11;

QY 2 TTDAATHVRGRSALLTSRLMTGNPSGRVEDDRLRLMDHARDGINCIDTADYGNR 61  
DB 59 tknameyrkiqdsdlnisevtumtfcgeqntekeshemysaieegincidtaeypip 118  
QY 62 LVR---GHTTELGRMLAOGGRRDPTLATKYGG--EMSERVNDG---LSARHIIAS 112  
DB 119 mkketqkidllyslswlk--sqgrdkivlatkvcysetsayirsdgellrvdaanlkes 176  
QY 113 CEGSLRRLGVDHIDVYOMHHIDR-----SAPWDEYWOAMDSLVSAGV 155

Db 177 vekskrtgtydlldlqhwprtyplfgdfyetskwrpsvpfiegqtrafdllveqkv 236  
Qy 156 SYVGSSNFGWHIAAQENARRHSLGMVSHOCLYNLAVR--HAELEVLPAAQAYGLGV 212  
Db 237 ryfgsnefsyvtcfvntakleglpklsiqngysllvrcryevdlvevchpkncvyl 296  
Qy 213 FAMSPLHGGLSGALEKTLAAG-TAVKSAO-----GRAQVLPBLSL-RPALEAVERKCRNLG 265  
Db 297 latsplggsgslsg--kylatdqeaeknarlnlfpymerykyslakeatigqvevakkyg 354  
Qy 266 EDPAEVGLAWYLSRPGIAGAVIGPTPEQLDSALKASAMT---LDEQALSELDEIF 318  
Db 355 ltpvelalgtvrdprpfvsttligatsvkqlkedidaflmterpfqevmadidavf 410  
RESULT 10  
AAG41676  
ID AAG41676 standard; Protein: 357 AA.  
AC AAG41676;  
XX 18-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51883.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hydrialsation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.

PR	05-AUG-1999;	9905-01471260;
PR	05-AUG-1999;	9905-0147303;
PR	06-AUG-1999;	9905-0147416;
PR	06-AUG-1999;	9905-0147419;
PR	09-AUG-1999;	9905-0147493;
PR	10-AUG-1999;	9905-01481771;
PR	11-AUG-1999;	9905-0148319;
PR	13-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148565;
PR	16-AUG-1999;	9905-0148684;
PR	17-AUG-1999;	9905-0149368;
PR	18-AUG-1999;	9905-01494175;
PR	20-AUG-1999;	9905-01494226;
PR	20-AUG-1999;	9905-01493722;
PR	20-AUG-1999;	9905-01493723;
PR	20-AUG-1999;	9905-01495929;
PR	23-AUG-1999;	9905-01499002;
PR	23-AUG-1999;	9905-0148930;
PR	25-AUG-1999;	9905-0150566;
PR	26-AUG-1999;	9905-0150884;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151303;
PR	31-AUG-1999;	9905-0151438;
PR	01-SEP-1999;	9905-0151930;
PR	07-SEP-1999;	9905-0152363;
PR	10-SEP-1999;	9905-0153070;
PR	13-SEP-1999;	9905-0153758;
PR	15-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154039;
PR	20-SEP-1999;	9905-0154779;
PR	22-SEP-1999;	9905-0155139;
PR	24-SEP-1999;	9905-0155486;
PR	24-SEP-1999;	9905-0155569;
PR	28-SEP-1999;	9905-0156584;
PR	28-SEP-1999;	9905-0156596;
PR	04-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157753;
PR	06-OCT-1999;	9905-0157865;
PR	07-OCT-1999;	9905-0158029;
PR	08-OCT-1999;	9905-0158232;
PR	12-OCT-1999;	9905-0158369;
PR	13-OCT-1999;	9905-0159293;
PR	13-OCT-1999;	9905-0159294;
PR	14-OCT-1999;	9905-0159329;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159331;
PR	14-OCT-1999;	9905-0159637;
PR	18-OCT-1999;	9905-0159638;
PR	21-OCT-1999;	9905-0160741;
PR	21-OCT-1999;	9905-0160767;
PR	22-OCT-1999;	9905-0160768;
PR	22-OCT-1999;	9905-0160981;
PR	22-OCT-1999;	9905-0160989;
PR	25-OCT-1999;	9905-0161404;
PR	25-OCT-1999;	9905-0161405;
PR	26-OCT-1999;	9905-0161359;
PR	26-OCT-1999;	9905-0161360;
PR	26-OCT-1999;	9905-0161361;
PR	28-OCT-1999;	9905-0161920;
PR	28-OCT-1999;	9905-0161922;
PR	29-OCT-1999;	9905-01621493;
PR	29-OCT-1999;	9905-01621492;

Query Match 20.1%; Score 347.5; DB 21; Length 357;  
Best Local Similarity 31.0%, Pred. No. 1.7e-26;  
Matches 108; Conservative 61; Mismatches 136; Indels 43; Gaps 11.

QY RGRSALTLSTRMLGCVNFSGRVEDDRLRLMDHARDRCINCLDTADMYGMFLYK---GH 66  
DB 10 :| | | | | : :| | | : : : : : : :| | | | | :| :| :|  
5 klgdgdlnlsetvmgtumtfcgeqntekeshemsyaleeglncldaeaayipmkkxetgk 64  
QY 67 TEELVGRWLAAOGGGRREDVTLLKYVG--EMSERVNDSG---LSARHLIASCESGLRRL 120  
DB 65 tdllysswtk--sqgrdkvlavtkvcygsersaylrtdsgellrvdaanlkvesvskrl 122  
QY 121 GVDHDIVYOMHIDR-----SAPWDEVMQMSDLVASGKSYVGSNNF 163  
DB 123 gfdydldllghwpdrtyrflfgdyfetskwpsvpfaeqrlrfqglivegkvrylgvsne 182  
QY 164 AGWHIAAQAENARRRHNSLGMVSHOCLYNLAVR---HAEEVLPAQAQVSLGVFAANSPLHG 220  
DB 183 tszygvtrefnrtckleglpkhivslngysllvrctryevdvlvercphncnvqllaysplgg 242  
QY 221 GLLSALEALKLAG-TAVKSAQ-----GRAQVLLPST-RPAIEAYEKFCNLGEDPAEVLG 273  
DB 243 gslsg-kyladgcatkmarinlfpgymerykgsjaketaeltqyevakkyygltprelal 300  
QY 274 AMVLSRPETAGAVIGPRPEQLDSALKASAMT---LDQALSELDEF 318  
DB 301 gfvrdrpfvtstcligatsvkqlkedafimterpfsgsvmadlavf 348

RESULT 11  
AAB79618  
ID AAB79618 standard; Protein: 312 AA.  
XX  
AC AAB79618;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:752.  
DX  
XX  
KW Corynebacterium glutamicum; carbon metabolism and energy production;  
KM SMP protein; sugar metabolism and oxidative phosphorylation protein;  
KM fine chemical production; organic acid; proteinogetic amino acid;  
KM nonproteinogetic amino acid; purine base; pyrimidine base; nucleoside;  
KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
KM carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
diagnosis; Corynebacterium diptheriae; evolutionary study.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200100844-A2.  
XX  
PD 04-JAN-2001.  
PF 23-JUN-2000; 2000MO-IB00943.  
XX  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031412.  
PR 08-JUL-1999; 99DE-1031413.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031431.  
PR 08-JUL-1999; 99DE-1031433.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031562.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1032180.  
PR 08-JUL-1999; 99DE-1032227.  
PR 08-JUL-1999; 99DE-1032230.  
PR 08-JUL-1999; 99DE-1032230.  
PR 08-JUL-1999; 99US-0145208.  
PR 14-JUL-1999; 99DE-1032924.



PR	31-AUG-1999;	99DE-1041370.	
PR	31-AUG-1999;	99DE-1041380.	
PR	31-AUG-1999;	99DE-1041394.	
PR	31-AUG-1999;	99DE-1041396.	
PR	03-SEP-1999;	99DE-1042076.	
PR	03-SEP-1999;	99DE-1042077.	
PR	03-SEP-1999;	99DE-1042079.	
PR	03-SEP-1999;	99DE-1042085.	
PR	03-SEP-1999;	99DE-1042087.	
PR	03-SEP-1999;	99DE-1042088.	
PR	03-SEP-1999;	99DE-1042095.	
PR	03-SEP-1999;	99DE-1042124.	
PR	03-SEP-1999;	99DE-1042129.	
PR	09-MAR-2000;	2000US-0187970.	
PA	(BADI ) BASF AG.		
XX			
XX	Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;		
XX	WPI; 2001-137957/14.		
DR	N-PSDB; AAF72186.		
XX			
XX	Nucleic acids from Corynebacterium glutamicum encoding metabolic		
PT	pathway proteins, useful for producing fine chemicals in		
PT	microorganisms, including organic acids, nonproteinogenic amino acids,		
PT	and purine and pyrimidine bases -		
PS	Claim 20; Page 1354-1355; 1737/p; English.		
XX			
CC	AAF1753 to AAF72330 encode the Corynebacterium glutamicum metabolic		
CC	pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum		
CC	MP nucleic acids are useful for the production of fine chemicals		
CC	in microorganisms, including organic acids, nonproteinogenic amino		
CC	acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,		
CC	saturated and unsaturated fatty acids, diols, carbohydrates, aromatic		
CC	compounds, vitamins, cofactors, polyketides and enzymes.		
SQ	Sequence 312.AA:		
Query Match	19.5%; Score 337; DB 22; Length 312;		
Best Local Similarity	30.3%; Pred. No.1.6e-25;		
Matches 96; Conservative 53; Mismatches 150; Indels 18; Gaps			
QY	11 LCRSALITRLMLGTVNESCGRVEDDALARLIMHARDG;INCIDTADMTGWRLYKHTTEL 70		
DB	2 YGSSGLIYRSLITGSTWGLEAAEGDIFKAFINSGLTIDVAPNY---tlgvaem 57		
QY	71 VRRMLAOGGGRREDYVLATKVGGE---MSEKVNDSGLSARHIIASCEGSLRRLGVDHD 126		
DB	58 lgtmI-daevrsasavvissagvmpalpgirrvd---cgrnlllqldvltlrainlntyid 113		
QY	127 VYQMHIDRSAPFWEQMDMSLVASGKSYVSSNFPAGMHIAA--AOENARRRSLGMV 184		
DB	114 lwsvgywdegtrphhevadltidvartgrvrygavrgysgwqlavthaasnaaasarpv 173		
QY	185 SHOCLYLNLAVRAAELEVPAAQAYGIGVANSPLNGLLSGLAEKTAAGTAVKSAQGRQ 244		
DB	174 vaqneysllieraaegellipatqlgygfiagapllqgyvltakyrseipdistaastgita 233		
QY	245 VILPST---RPAIAYEYKFCRNLEDPEVGLAWLSPGJAGAVIGPRTPEQDLSALK 300		
DB	234 evqsyidnqgrilvaldaaakgylgispavattctwrdipgytaviavgartheqishlik 293		
QY	301 ASAMTIDEGALSELDEI 317		
DB	294 aevstlpcpitqalddv 310		
RESULT 13			
FAW23389			
AAW23389	standard; protein; 329 AA.		
XX			

AC	AAW2389,
XX	
DT	02-APR-1998 (first entry)
DE	Shaker-like potassium ion channel beta-subunit core region Kv beta 2.
XX	
KM	Shaker-like potassium ion channel; SPC: beta subunit; core region;
KV	Kv beta 2; N-terminal A and B box; NAB; treatment; cardiac disease;
KW	tumour; auto immune disease.
XX	
OS	Mammalia.
XX	
PN	WO9731112-A2.
PD	28-AUG-1997.
XX	
PF	18-FEB-1997; 97WO-US02292.
XX	
PR	23-FEB-1996; 96US-0606143.
XX	
PA	(UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
Ll M;	
DR	WPI: 1997-435164/40.
XX	
Pt	Polypeptide(s) derived from Shaker-like potassium ion channel alpha
PT	and beta subunits - used to alter potassium ion levels in a cell,
Pr	e.g. for treating neurological disorders, tumours, metabolic disease
XX	
PS	Claim 15; Page 53; 106pp; English.
CC	This polypeptide sequence Kv beta 2 consists of the core region of a
CC	beta-subunit of a Shaker-like potassium ion channel (SPC). This core
CC	region can bind to a polypeptide consisting of the N-terminal A and B
CC	box (NAB) domain and the NAB-SI (the first transmembrane spanning domain)
CC	linking region of the alpha-subunit of SPC. The polypeptides or the
CC	nucleic acid encoding them can be introduced into the cytoplasm of a
CC	cell to modulate the flow of potassium ions through a cytoplasmic cell
CC	membrane. Potassium ion channels regulate the action potentials, cardiac
CC	pacemaking and neurotransmitter release in excitable tissues. In
CC	non-excitable tissues they play important roles in hormone secretion,
CC	cell proliferation, cell volume regulation and lymphocyte
CC	differentiation. Molecules which bind to the alpha or beta-subunit
CC	polypeptides can be detected by contacting the polypeptides, with a
CC	putative NAB and NAB-SI linking region of an alpha-subunit or with a
CC	putative core region of a beta-subunit respectively, and determining
CC	whether or not binding occurs. These polypeptides and the encoding
CC	nucleic acids may be useful in the treatment of a huge variety of
CC	disorders, e.g. neurological disorders, tumours, metabolic diseases,
CC	cardiac disease and autoimmune disease.
XX	
Sequence	329 AA:
Query Match	19.2%; Score 331; DB 18; Length 329;
Best Local Similarity	28.6%; Pred. No. 6.8e-25;
Matches 94; Conservative 60; Mismatches 137; Indels 38; Gaps 9;	
OY	11 LGRSALLISRLMLGT-VNFSGVEYEDDALARLMDHARDGINCILDPADMGKRLYKGHTEE 69
Db	4 lqsklyrvcslglfwtvtfgqldemeahlnltlaydnginfdaeyaa---agkaev 59
OY	70 LVGRMLAOGGGRRPDTLVATKY--GGEMSERNDGSLSRHHITASEGSIRLRGVHDIDV 127
Db	60 vlgnlkikkwytsslvltklifwgkae---terglsrkhlleglkaslerlqlleyavh 116
OY	128 YOMNHIDRSAPDEWQAOANDSLVASGKYSYYGSSNFPAGWHIAAQENARRHSLGMVSHQ 187
Db	117 vtanrpdrptmceelvrantbhvingmamwytsrtswsmelineaysvatqflippiceq 176
OY	188 CLYNLAVRH-AALEVLPAQAQGLGFANWSPLHGGLLSGALEKLAAG----- 233

```

Db 177 aeyhmfgrekeveqjpebfhkiyvgamtwspiacqlvsg---kydsq:ppyrasakkyq 233
QY 234 ---TWKSAQGAQVLLPSLRPAIEAYEKEFCRNLDGPAEVLAWLSRPGIAGAVIGPR 290
Db 234 wlkdkllseegrtq-----gakikelaerlgtclpqlalawclmrnegssvllgas 287
QY 291 TPEQDLSALKASAM--TLDEQALSELDEL 317
Db 288 naeqimenaigaqlvlpklssisvheidst 316

```

# RESULT 14 AAM69711

ID AAM69711 standard; Protein; 344 AA.

AC AAM69711;

DT 26-OCT-1998 (first entry)

DE Streptomyces clavuligerus protein sequence of orf1p1.

KW Streptomyces clavuligerus; bacterial gene; clavulanic acid; 5S clavum;  
5R clavum; cas1; ORF; open reading frame; biosynthesis.

OS Streptomyces clavuligerus.

PN WO9833896-A2.

PD 06-AUG-1998.

PE 02-FEB-1998; 98WO-EP00644.

PR 04-FEB-1997; 97GB-0002218.

PA (SMIK ) SMITHKLINE BEECHAM PLC.  
(UYAL-) UNIV ALBERTA.

PI Anders C, Barton B, Griffin JP, Jensen S, Mosher RH;  
Paradkar AS;

DR WPI: 1998-437451/37.

DR N-PSDB: AAV50486.

PT DNA comprising defective 5S clavain biosynthesis gene(s) from  
Streptomyces clavam - useful for producing clavulanic acid without

PT production of 5S clavam or clavam-2-carboxylates

PS Claim 3; Page 15-24; 29pp; English.

XX The present sequence represents the protein sequence of a gene specific  
CC for 5S clavam biosynthesis from Streptomyces clavuligerus (SC) and which  
CC is not essential for 5R clavam biosynthesis. The present invention also  
CC describes: (1) a process for improving 5R clavam production in a  
CC suitable microorganism comprising manipulation of DNA encoding as above  
CC and its inclusion in the microorganism; (2) a process for improving 5R  
CC clavam production in SC comprising disrupting or otherwise making  
CC defective DNA regions flanking cas 1; (3) a process for the  
CC identification of a microorganism suitable for high 5R clavam production  
CC comprising a preliminary screening for microorganisms with low or no 5S  
CC clavam production; (4) a microorganism which is capable of 5R clavam  
CC production and low or no 5S clavam production obtainable by a process as  
CC in (3); (5) clavulanic acid (CA) obtainable by the fermentation of a  
CC microorganism as in (4), and (6) CA which is free of any 5S clavam; (7)  
CC CA which is free of any clavam-2-carboxylate. The methods and products  
CC can be used to produce organisms capable of producing increased amounts  
CC of clavams suitably e.g. CA, for use as antibiotics. The methods can  
CC also be used for the production of CA without the production of 5S  
CC clavam or clavam-2-carboxylate.

XX Sequence 344 AA;

Query Match 19.0%; Score 329; DB 19; Length 344;  
Best Local Similarity 33.4%; Pred. No. 1.1e-24;  
Matches 105; Conservative 45; Mismatches 132; Indels 32; Gaps 8;

```

QY 23 LGTVNSGHEVEDDALLRMDHARDGINCILDTADMYGMRLYGHNTELVGRMLAOGGRR 82
Db 38 lpttdygrprdaratratatlaavadgvtllldtadvg-----lgagellgrava---grr 90
QY 83 EDTVLATKVGEMSERVNDGLSAR--HIIASCESGLRRLGVHDIDVYOMHIDRSAPMD 140
Db 91 devliatkfgmvrssdgaagqilcgpsyrracercslrrlgrtdridlyghwtdpavple 150
QY 141 EYWMAMDLSVAGSKSYVSSNFAGHNHIAADENARRRSLGVSQCIXNLAVRIAELE 200
Db 151 etvgavaelvregekvyrrlglsps---aalrradavhpvtav--qsewslwsglde 204
QY 201 VLPAQAQYGLGVAFVMSPLHGLLSCAL-----EKLAAGTAVKSAQ--GRAQVLLPSL 250
Db 205 vvpvcrelglgivaayapigrftlgrtdldgdedfrrgqprfapalarnrllhrl 264
QY 251 RPAIEAYEKEFCRNLDGPAEVLAWLSRPGIAGAVIGPRPEQDLSALKASAMTLDEQA 310
Db 265 rpvad-----glgltaqlawlhrrgedvvpigfcanpahladnlaasirldtrs 317
QY 311 LSELDELFPFAVSG 324
Db 318 laevtaaishpvsq 331

```

# RESULT 15

ID AAB08255 standard; Protein; 344 AA.

AC AAB08255;

DT 04-DEC-2000 (first entry)

DE The 6C5 cell wall antigen of Candida albicans.

KW Cell wall protein; yeast infection; pathogenesis inhibitor; 6C5 antigen;  
pathogenic yeast.

OS Candida albicans.

PN WO200048620-A1.

PD 24-AUG-2000.

PE 18-FEB-2000; 2000WO-US04228.

PR 19-FEB-1999; 99US-0120764.

PR 19-FEB-1999; 99US-0120765.

PA (UYVI-) UNIV VIRGINIA PATENT FOUND.

PI Hazen KC, Singleton DR, Masouka J, Wu JG, Glee PM;

DR WPI: 2000-565334/52.

DR N-PSDB: AAA63930.

PT New isolated hydrophobic protein antigens derived from the cell walls  
of Candida albicans and their corresponding antibodies, useful for

PT detecting and treating yeast infection

PS Disclosure; Fig 5A-C; 62pp; English.

XX The present sequence represents a cell wall protein, designated 6C5, of  
CC Candida albicans. The cell wall peptides and their corresponding  
CC antibodies are useful for treating yeast infections in a patient,  
CC as well as for preventing or detecting infection of any yeast  
CC type. These include opportunistic and nosocomial infections. The  
CC peptides are also useful in the development of peptidomimetics and/or  
CC small organic molecules with therapeutic, diagnostic and/or research

CC use. The antibodies are effective inhibitors of hydrophobically mediated  
 CC pathogenesis events. They are also useful in diagnostic tests  
 CC (i.e. detect, diagnose and serotype yeast infection), and for detecting  
 CC the presence of Candida and other pathogenic yeasts, in a biological  
 CC sample. Furthermore, they may be used as research and development tools.

XX  
 SQ Sequence 344 AA:

Query Match 18.6%; Score 321; DB 21; Length 344;  
 Best Local Similarity 27.0%; Pred. No. 7.3e-24;  
 Matches 93; Conservative 66; Mismatches 152; Indels 34; Gaps 8;

```

QY 1 MTTDAATHV-RLGRSALLTSLMLGV-----NFSGRVEDDALRLMDHARDGICNL 52
   | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 msidsksmvrlrlgskglkvntvaavgtmrlygswrgfngdi--deciklikfcyngfrtf 58

QY 53 DTADMTGWRLYKHTPELVGRMLAOGGGRREDVYLATKVGEMSERVNDSE----- 102
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 dtadctys---ngkseellglfllkkyntpreivlltkcyfsvkdaedssleldpldym 114

QY 103 ---GLSARHIIASCESGLRRLGVHDIDVYOMHHIDRSAPWDEWQAMDSLVASGKVSYG 159
   ||| : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 115 ngkylsrkhllaaasvkrlygt-yldvlglnldhevtveymrslndvvegglarylg 173

QY 160 SSNFAGWHIAAQBENARRHSLGNVSHQCCLYNLAVRHAELEVLPAQAQYGLGVFAWSPLH 219
   : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 174 assmktwefvelgnvakangwhgflsmqshysllyrederelndycknglglipwspng 233

QY 220 GGLLSGALKELAGTAVKSAOGRQAVLLPSLRPA---IEAYEKFCNLGDEPAVEGLAW 275
   ||| : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 234 ggvlycrpfdscktkqfidnkqwsllglenvrdackliivdrveelsvkynasmmqvs law 293

QY 276 VLSRPGIAGAVIGPRTPEQLDSALKASAMTLDQALSELDEIFPA 320
   : : : | : | : | : | : | : | : | : | : | : | : | : |
Db 294 ciakgyi--plagvskfegaeelvgifkvnlteddkyleepgha 336

```

Search completed: June 22, 2001, 09:24:33  
 Job time: 109 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:22:59 ; Search time 13.07 Seconds  
(without alignments)  
513.249 Million cell updates/sec

Title: US-09-463-705A-2

Perfect score: 1728  
Sequence: 1 MTDAATHVRLGRSALLTSR.....IDETFPVAVSGGAPEAWLQ 333

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgcn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgcn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgcn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgcn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgcn2\_6/ptodata/2/1aa/PCRTUS\_COMB.pep:\*  
6: /cgcn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	1114	2	US-08-576-626A-31 Sequence 31, Appl
2	356	20.6	329	2	US-08-606-143-1 Sequence 1, Appl
3	356	20.6	329	2	US-08-606-143-1 Sequence 3, Appl
4	331	19.2	329	2	US-08-606-143-2 Sequence 2, Appl
5	208.5	12.1	331	2	US-08-907-674-1 Sequence 1, Appl
6	208.5	12.1	331	2	US-09-215-087-1 Sequence 1, Appl
7	208.5	12.1	331	2	US-09-391-959-1 Sequence 1, Appl
8	196.5	11.4	327	3	US-08-907-674-3 Sequence 3, Appl
9	196.5	11.4	327	3	US-09-215-087-3 Sequence 3, Appl
10	196.5	11.4	327	3	US-09-391-959-3 Sequence 3, Appl
11	177	10.2	278	1	US-07-941-414-1 Sequence 1, Appl
12	177	10.2	278	1	US-08-249-377A-1 Sequence 1, Appl
13	177	10.2	278	1	US-08-585-595-1 Sequence 1, Appl
14	177	10.2	278	1	US-08-749-337-1 Sequence 1, Appl
15	161.5	9.3	318	2	US-08-336-198C-3 Sequence 3, Appl
16	157	9.1	277	1	US-08-585-595-2 Sequence 2, Appl
17	149	8.6	310	4	US-09-166-412-4 Sequence 4, Appl
18	140	8.1	309	4	US-09-166-412-2 Sequence 2, Appl
19	111.5	6.5	316	3	US-08-801-344-4 Sequence 4, Appl
20	108	6.2	323	3	US-08-853-839-2 Sequence 2, Appl
21	104	6.0	323	4	US-08-532-896-2 Sequence 2, Appl
22	99	5.7	316	1	US-08-585-595-3 Sequence 3, Appl
23	99	5.7	3170	2	US-07-642-734C-5 Sequence 5, Appl
24	97.5	5.6	3170	4	US-08-439-009A-5 Sequence 5, Appl
25	96.5	5.6	3567	2	US-07-642-734C-4 Sequence 4, Appl
26	96.5	5.6	3567	2	US-08-439-009A-4 Sequence 4, Appl

28	95.5	5.5	7257	4	US-09-335-409-5	Sequence 5, Appl
29	93	5.4	431	1	US-08-391-939-18	Sequence 18, Appl
30	93	5.4	431	1	US-08-484-274A-18	Sequence 18, Appl
31	92.5	5.4	2257	1	US-08-611-107-10	Sequence 10, Appl
32	92.5	5.4	2257	2	US-08-422-560A-10	Sequence 10, Appl
33	92.5	5.4	2257	4	US-08-468-793-10	Sequence 10, Appl
34	92	5.3	1864	2	US-08-804-227C-3	Sequence 3, Appl
35	91.5	5.3	569	2	US-08-467-822-27	Sequence 27, Appl
36	91	5.3	4551	4	US-09-320-878-1	Sequence 1, Appl
37	89	5.2	407	2	US-08-926-327-2	Sequence 2, Appl
38	89	5.2	407	3	US-09-119-918-2	Sequence 2, Appl
39	88.5	5.1	402	2	US-08-403-852D-19	Sequence 19, Appl
40	88.5	5.1	402	3	US-08-510-646B-20	Sequence 20, Appl
41	88.5	5.1	402	4	US-09-231-818-19	Sequence 19, Appl
42	88.5	5.1	430	1	US-08-391-939-5	Sequence 5, Appl
43	88.5	5.1	430	1	US-08-484-274A-5	Sequence 5, Appl
44	87.5	5.1	1611	2	US-08-804-227C-5	Sequence 5, Appl
45	87.5	5.1	4472	2	US-08-804-227C-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-576-626A-31  
Sequence 31, Application US/08576626A  
Patent No. 5998194  
GENERAL INFORMATION:  
APPLICANT: Summers, R.G.  
APPLICANT: Katz, L.  
APPLICANT: Donadio, S.  
APPLICANT: Staver, M.J.  
TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
TITLE OF INVENTION: BIOSYNTHESIS GENES  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/576,626A  
FILING DATE: 21-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dianne Casuco  
REGISTRATION NUMBER: P-40,943  
REFERENCE/DOCKET NUMBER: 5857.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (847) 938-3137  
TELEFAX: (847) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1114 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5998194e  
US-08-576-626A-31  
Query Match 100.0%; Score 1728; DB 2; Length 1114;

Best Local Similarity 100.0%; Pred. No. 1.9e-178;  
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTAAATHVRLGRSALTSLRLMLGTVNFSGRVEDDRLRLMDHARDGINCIDPTADMYGW 60  
|||  
Db 782 MTTDAATHVRLGRSALTSLRLMLGTVNFSGRVEDDRLRLMDHARDGINCIDPTADMYGW 841  
QY 61 RLKGHTEELVGRWLAOGGGRREDTVALTKVGGMSERVDLSARHIIASCBSLRRL 120  
|||  
Db 842 RLKGHTEELVGRWLAOGGGRREDTVALTKVGGMSERVDLSARHIIASCBSLRRL 901  
QY 121 GVHDIDYVOMHHIDRSAPMDEVMQAMDSIVASGKVSIVGSSNFGWHIAAQENAAARRHS 180  
|||  
Db 902 GVHDIDYVOMHHIDRSAPMDEVMQAMDSIVASGKVSIVGSSNFGWHIAAQENAAARRHS 961  
QY 181 LGMVSHOCLYNLAVRHALEVLPAQAQVGLGVFAMSPDLHGGLSGALEKLAGAIVKSAQ 240  
|||  
Db 962 LGMVSHOCLYNLAVRHALEVLPAQAQVGLGVFAMSPDLHGGLSGALEKLAGAIVKSAQ 1021  
QY 241 GRAOVLPLSLPAIEAYEKFCRNIGEDPAEYGLAWLSRPGIAGAVIGPRTPEQLDSALK 300  
|||  
Db 1022 GRAOVLPLSLPAIEAYEKFCRNIGEDPAEYGLAWLSRPGIAGAVIGPRTPEQLDSALK 1081  
QY 301 ASANTLDEQALSELDETFPAVASGAAPFAMLO 333  
|||  
Db 1082 ASANTLDEQALSELDETFPAVASGAAPFAMLO 1114

## RESULT 2

US-08-606-143-1  
; Sequence 1, Application US/08606143  
; Patent No. 5856155  
; GENERAL INFORMATION:  
; APPLICANT: Li, Min  
; TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR  
; TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/606,143  
; FILING DATE: 23-FEB-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kilyk Jr., John  
; REGISTRATION NUMBER: 30763  
; REFERENCE/DOCKET NUMBER: 71756  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 329 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-606-143-1

Query Match 20.6%; Score 356; DB 2; Length 329;

Best Local Similarity 30.9%; Pred. No. 2.1e-30;  
Matches 102; Conservative 57; Mismatches 131; Indels 40; Gaps 10;

QY 11 LGRSALTSLRLMLGT-VNFGSRVEDDRLRLMDHARDGINCIDPTADMYGRLYKHTTE 69  
|||  
Db 4 LKSGKLVSCIGLSTWTFGGQISDEVAERLMTIAYSSGVNLPDTAVYA---AGRAEV 59  
QY 70 LVGRWLAOGGGRREDTVALTKV--GGMSESRVDLSARHIIASCBSLRRLGVHDIDV 127  
|||  
Db 60 ILGSIIRKKGWRSSSLVITTKRLYWGKAE---TERGSRHIIIEGLGSLQRLQLEVVDV 116  
QY 128 YOMHHIDRSAPMDEVMQAMDSIVASGKVSIVGSSNFGWHIAAQENAAARRHSIGVSHQ 187  
|||  
Db 117 VFARPPSNTPMEIVIRAMTHVINQGMAMWGTSRWSAMEIMEYVSARQFMNIPVCEQ 176  
QY 188 CLYNLAVRH-ALEVLPAQAQVGLGVFAMSPDLHGGLSGALEKLAGAIVKSAQ----- 234  
|||  
Db 177 AEYHLFQREKVEYVLPPELRYKIGVAMTWSPFLACGIIISG---KIGKVPRESSRASLQCY 233  
QY 235 ---AVKSAQGRAQV-LPLSLPAIEAYEKFCRNIGEDPAEYGLAWLSRPGIAGAVIGP 289  
|||  
Db 234 WLKERVISEGRKQNKTKDLSPIAE-----RLGCTLPQLAVAMCLRNQGVSVLLGS 286  
QY 290 RTPEQLDSALKASAM--TLDEQALSELDEI 317  
|||  
Db 287 STPEQLIENLGAIOVLPRMTSHVNEIDNI 316

## RESULT 3

US-08-606-143-3  
; Sequence 3, Application US/08606143  
; Patent No. 5856155  
; GENERAL INFORMATION:  
; APPLICANT: Li, Min  
; TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR  
; TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/606,143  
; FILING DATE: 23-FEB-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kilyk Jr., John  
; REGISTRATION NUMBER: 30763  
; REFERENCE/DOCKET NUMBER: 71756  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 329 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-606-143-3

Query Match 20.6%; Score 356; DB 2; Length 329;

Best Local Similarity 30.9%; Pred. No. 2,1e-30;  
Matches 102; Conservative 57; Mismatches 131; Indels 40; Gaps 10;

QY 11 LGRSALLTSRLMGT-VNFSGRVEDDALARMDHARDGINCIDTADMYGWRLYKGTTEE 69  
DB 4 LKSGRLVSCLGIGTWTGCGQIDSEVAERLMTIAGESVNFEDTAEVYA----AGKAV 59  
QY 70 LVGRMLAOGGGRREDTVALTKV--GGEMSERVNDGSLARHIIASCEGSLRILGVHDIV 127  
DB 60 ILGSIKKKGMRRSSLVITTKLYWGKAE---TERLSKRHIIEGKSLRLQLELYVDV 116  
QY 128 YOMHIDRSAPDEWQANDSLVASKSVYGSNPNAGHIIAAGQNAARRISLGVSHQ 187  
DB 117 VANRRDPTPMEETVRAKTHVINOAMWGTSRWSMEIEMAYSVARQFNILPPICEQ 176  
QY 188 CLYNLAVRH-AELEVLPAQAAYGLGVFAMSPHLGGLSGALEKLAGT----- 234  
DB 177 AEYHMFQRKEVQVQPELPHKIGVGAMTWSPLACGIVSG---KYSGIPYRSALIKYQ 233  
QY 235 ---AVKSAQRAQV-LPSPRAIEAYEKFGRNIGEDPAEVLAVLSRPGIAGAVIGP 289  
DB 234 WKEKRVSEGRROQNKILDSPIAE-----RLGCTLPOLAVMCLNREGVSVLLGS 286  
QY 290 TPEQDLSALKASAM--TIDEQALSEIDEI 317  
DB 287 STEPOLIENLGAIQVLPKMTSHVNEIDNI 316

RESULT 4  
US-08-606-143-2

; Sequence 2, Application US/08606143  
; Patent No. 5856155  
; GENERAL INFORMATION:  
; APPLICANT: LI, Min  
; TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR  
; TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/606,143  
; FILING DATE: 23-FEB-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kilyk Jr., John  
; REGISTRATION NUMBER: 30763  
; REFERENCE/DOCKET NUMBER: 71756  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 329 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-606-143-2

Query Match

19.2%; Score 331; DB 2; Length 329;

Best Local Similarity 28.6%; Pred. No. 1,1e-27;  
Matches 94; Conservative 60; Mismatches 137; Indels 38; Gaps 9;

QY 11 LGRSALLTSRLMGT-VNFSGRVEDDALARMDHARDGINCIDTADMYGWRLYKGTTEE 69  
DB 4 LKSGRLVSCLGIGTWTGCGQITDEMAMHMLMTIADNGLNFTDAEVYA----AKKAV 59  
QY 70 LVGRMLAOGGGRREDTVALTKV--GGEMSERVNDGSLARHIIASCEGSLRILGVHDIV 127  
DB 60 VLGNIKKKGMRRSSLVITTKIFWGKAE---TERLSKRHIIEGKSLERLQLELYVDV 116  
QY 128 YOMHIDRSAPDEWQANDSLVASKSVYGSNPNAGHIIAAGQNAARRISLGVSHQ 187  
DB 117 VANRRDPTPMEETVRAKTHVINOAMWGTSRWSMEIEMAYSVARQFNILPPICEQ 176  
QY 188 CLYNLAVRH-AELEVLPAQAAYGLGVFAMSPHLGGLSGALEKLAGT----- 233  
DB 177 AEYHMFQRKEVQVQPELPHKIGVGAMTWSPLACGIVSG---KYSGIPYRSALIKYQ 233  
QY 234 ---TAVKSAQRAQVLLPSPRAIEAYEKFGRNIGEDPAEVLAVLSRPGIAGAVIGP 290  
DB 234 WKEKRVSEGRROQNKILDSPIAE-----QAKLKLQAIARLGCTLPOLAVMCLNREGVSVLLGS 287  
QY 291 TPEQDLSALKASAM--TIDEQALSEIDEI 317  
DB 288 NAEQLEMENIGAIQVLPKLSSTVHEIDSI 316

RESULT 5  
US-08-907-674-1

; Sequence 1, Application US/08907674  
; Patent No. 5919685  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/907,674  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0362 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:

LIBRARY: BRAINOT14  
CLONE: 1596452  
US-08-907-674-1

Query Match 12.1%; Score 208.5; DB 2; Length 331;  
Best Local Similarity 23.9%; Pred. No. 2.2e-14;  
Matches 83; Conservative 49; Mismatches 135; Indels 81; Gaps 10;

QY 23 LGTVNFGSRVEDDALRLMDHARDGNCIDTADMGWRMLYKGTETELVGRMLAOGGRR 82  
14 LGAMKGRMDAPTSAAVTAFLEHGTETDFAVIS---EGQSETILGGLRLGGGD 69  
QY 83 EDVVLATKVGEMSERVNDGSLRNHIIASCEGSLRLRGVDHIDVYOMNHIDRSAPWDEV 142  
70 CRVKIATK-----ANPMDGSLKPDVSRSQLETSIKRLQCPQVDLFYLHTRPDHGTPEET 124  
QY 143 WQAMDSLVASGKSYVSSNPFAGNHIAAOENARRHSLGMV---SHOCLYNLAVRHAEL 199  
125 LHACQRLHOGKFEVELGLSNVYASWEVA---EICTLCKSNGMILPTVYQGMYNATTRQVET 181  
QY 200 EVLPAAQVGLGVFAMSPRLGGLLSGAL-----EKL 230  
182 ELFPCLRHFGRLRYAIVPLAGLITGKYKEDKQKOPVGRFGNTWAEMRYRNRYKEHH 241  
QY 231 AAGTAV--KSAQGRAOVLPSLRPAIEAYEKFGRNLGEDPAEVLAVLSRPGIAGA--- 285  
242 FEGIALVEKALQAAVYASAPSVTSA-----ALRMVYHHSQLOGAHGD 283  
QY 286 --VIGPRTPEQDLSALKASAMTDEQALSELDEFPAVASGAAPRW 331  
284 AVILGMSLEQLERONT-----AATEGPLEPAVVD--AFNQAM 319

## RESULT 6

US-09-215-087-1  
Sequence 1, Application US/09215087  
Patent No. 5981244

GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/215,087  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/907,674  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0362 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 331 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAINOT14  
CLONE: 1596452  
US-09-215-087-1

Query Match 12.1%; Score 208.5; DB 2; Length 331;  
Best Local Similarity 23.9%; Pred. No. 2.2e-14;  
Matches 83; Conservative 49; Mismatches 135; Indels 81; Gaps 10;

QY 23 LGTVNFGSRVEDDALRLMDHARDGNCIDTADMGWRMLYKGTETELVGRMLAOGGRR 82  
14 LGAMKGRMDAPTSAAVTAFLEHGTETDFAVIS---EGQSETILGGLRLGGGD 69  
QY 83 EDVVLATKVGEMSERVNDGSLRNHIIASCEGSLRLRGVDHIDVYOMNHIDRSAPWDEV 142  
70 CRVKIATK-----ANPMDGSLKPDVSRSQLETSIKRLQCPQVDLFYLHTRPDHGTPEET 124  
QY 143 WQAMDSLVASGKSYVSSNPFAGNHIAAOENARRHSLGMV---SHOCLYNLAVRHAEL 199  
125 LHACQRLHOGKFEVELGLSNVYASWEVA---EICTLCKSNGMILPTVYQGMYNATTRQVET 181  
QY 200 EVLPAAQVGLGVFAMSPRLGGLLSGAL-----EKL 230  
182 ELFPCLRHFGRLRYAIVPLAGLITGKYKEDKQKOPVGRFGNTWAEMRYRNRYKEHH 241  
QY 231 AAGTAV--KSAQGRAOVLPSLRPAIEAYEKFGRNLGEDPAEVLAVLSRPGIAGA--- 285  
242 FEGIALVEKALQAAVYASAPSVTSA-----ALRMVYHHSQLOGAHGD 283  
QY 286 --VIGPRTPEQDLSALKASAMTDEQALSELDEFPAVASGAAPRW 331  
284 AVILGMSLEQLERONT-----AATEGPLEPAVVD--AFNQAM 319

## RESULT 7

US-09-391-959-1  
Sequence 1, Application US/09391959  
Patent No. 6071704

GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Shah, Purvi  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/391,959  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/907,674  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

```

; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT14
; CLONE: 1596452
;
; US-09-391-959-1

```

```

Query Match      12.1%; Score 208.5; DB 3; Length 331;
Best Local Similarity 23.9%; Pred. No. 2.2e-14;
Matches 83; Conservative 49; Mismatches 135; Indels 81; Gaps 10;

```

```

QY 23 LGTVNSGVEVDDDLRLMDHARDGINCIDTADMYGWLKGTTELVGRWLAOGGRR 82
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 14 LGAMEGRMDAPTSAATRAFLERHTEIDTAFVYS---EGQSEITLIGLGLRLGGG 69
QY 83 EDTVLTAKYGGEMSERVNDGSLARHIIASCGSLRLGLVDHIDVYQMHIDRSAPWDE 142
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 70 CRYATKTK---ANPMDKSLKPDSEVRSQLETSKRLQCPQVDLYLTPHGTVEET 124
QY 143 WQAMDSLVAAGKSYVSSNFAGWHIAAOENARHSLGMV---SHOCLYNLAVRHAEL 199
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 125 LIAQGLHDEGKVELGLSNYSWEVA---EICTLCKSNQWIIPTYYQGYNAITQVET 181
QY 200 EYLPAQAAYGLGVFANSPHGLLSGAL-----EKL 230
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 182 ELFPCLRHGRLFRYAYNPAGLLTGKYEKDKQPVGRFGNTWAEMRYRMYKEHH 241
QY 231 AGTAV--KSAQRAVLLPSLRPALEAYEKFCRNIGEDPAEGLAWLSRPIAQA--- 285
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 242 FEELALVEALQAAYASAPSVTA-----ALWMMHNSQLQAGHD 283
QY 286 --VIGRPEQLDSALKASAMTLEQATSELEIFPAPVAVASGAAPPAW 331
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 284 AVILGMSLEQLLEONL-----AATEEGLEPAVVD--AFNQAW 319

```

```

RESULT 8
US-08-907-674-3
; Sequence 3, Application US/08907674
; Patent No. 5919685
;
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN BI ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/907,674
; FILING DATE: Herewith
; CLASSIFICATION: 514

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 433611
;
; US-08-907-674-3

```

```

Query Match      11.4%; Score 196.5; DB 2; Length 327;
Best Local Similarity 23.5%; Pred. No. 4.2e-13;
Matches 80; Conservative 49; Mismatches 146; Indels 65; Gaps 12;

```

```

QY 23 LGTVNSGVEVDDDLRLMDHARDGINCIDTADMYGWLKGTTELVGRWLAOGGRR 82
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 10 LGAMEGRMDAPTSSASVAFQGRHTEIDTAFVYA---NQSEITLIGD-LGLIGRS 64
QY 83 EDTV-LATKVGEMSERVNDGSLARHIIASCGSLRLGLVDHIDVYQMHIDRSAPWDE 141
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 65 GCRVKATRAAPMFQGT---LKPADVRLQLETSKRLQCPVDLYLTPHGTPIEE 119
QY 142 WQAMDSLVAAGKSYVSSNFAGWHIAAOENARHSLGMVSHOCLYNLAVRHAELV 201
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 120 TLQACHVHDEGKVELGLSNYSWEVAELCTLCKNGWIMPTVYQGMNATITQVETEL 179
QY 202 LPAQAAYGLGVFANSPHGLLSGALAKLAAGTAVKSAQRAVLLPSLR---PAIEAV 257
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 180 FPLRLHFGRLFRYAFNPLAGLLTGRYK-----YQDKQDN---PESRFQNPFSQLY 228
QY 258 -----EKFCRNIGEDPAEGLAWLSRPIAGAVIGRTPQLDSALK----- 300
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 229 MDRYKKEHF-----NGTALV---EKALKTTYGTPAPSMISAAYRMVYHNSQLG 275
QY 301 --ASAMTLEQALSELDEIFPAPVAVASGAAP-----EAW 331
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 276 TQGDVAIIGMSLEQLLEONLALVEEGPLEPAVVDARFQAW 315

```

```

RESULT 9
US-09-215-087-3
; Sequence 3, Application US/09215087
; Patent No. 5981244
;
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN BI ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,087
; FILING DATE: Herewith
; CLASSIFICATION: 514

```

```
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,087
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/907,674
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 433611
US-09-215-087-3
```

```
Query Match 11.48; Score 196.5; DB 2; Length 327;
Best Local Similarity 23.58; Pred. No. 4,2e-13;
Matches 80; Conservative 49; Mismatches 146; Indels 65; Gaps 12;
```

```
QY 23 LGTVNFSGRVEDDRLIMHARDGINCLDTADMYGMRLYKGTTELVGFWLAOGGRR 82
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 10 LGAMEMGRNDVYSSASVRFQRGHTEIDTAHYA-----NGOSEITLGD-LGLGGRS 64
QY 83 EDIV-LATKYGGENSERVNDGSLARHIIASCEGSLRLGVDHIDVYOMHHIDRSAPWDE 141
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 65 GCKVKIATKAPMGKGT-----LKPADVRFQLETSKRLQCPRYDLFYLHPDHGTFIEE 119
QY 142 VMQAMDSLAVSGKSYVSSNFAGMHITAAQENARHSLGMSHOCILYNLAVNHALEEV 201
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 120 TLQCHHVHOGKFEVLGLSNVSWVEAEICTLCKKNGWIMPTVYOGMYNAITRQVETEL 179
QY 202 LPAQAYGLGVFAMSPILHGLLSGALKLAAGTAVKSAOGRAGVILPSLR-----PAIEAY 257
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 180 FPCLRHFGRLRYAFNPLAGLITGRK-----YQKDKGN-----PESRFGCPFSQLY 228
QY 258 -----EKFCRNIGEDPAEYGLAWLSRPGIAGAVIGPRTPEQLDSALK----- 300
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 229 MDRYWKKEHF-----NGIALV---EKALKTYGPTAPSMISAAYVWMYHHSQLKG 275
QY 301 --ASAMTLDEQALSELDEIFPAVASGGAAP-----EAW 331
DB 276 TQGDVAILGMSSTLEQLEONLALVEEGPLEPAVVADAFDOAW 315
```

```
RESULT 10
US-09-391-959-3
Sequence 3, Application US/09391959
Patent No. 6071704
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Puryi
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
```

```
STATE: CA
COUNTRY: USA
Zip: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/391,959
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/907,674
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 433611
US-09-391-959-3
```

```
Query Match 11.48; Score 196.5; DB 3; Length 327;
Best Local Similarity 23.58; Pred. No. 4,2e-13;
Matches 80; Conservative 49; Mismatches 146; Indels 65; Gaps 12;
```

```
QY 23 LGTVNFSGRVEDDRLIMHARDGINCLDTADMYGMRLYKGTTELVGFWLAOGGRR 82
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 10 LGAMEMGRNDVYSSASVRFQRGHTEIDTAHYA-----NGOSEITLGD-LGLGGRS 64
QY 83 EDIV-LATKYGGENSERVNDGSLARHIIASCEGSLRLGVDHIDVYOMHHIDRSAPWDE 141
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 65 GCKVKIATKAPMGKGT-----LKPADVRFQLETSKRLQCPRYDLFYLHPDHGTFIEE 119
QY 142 VMQAMDSLAVSGKSYVSSNFAGMHITAAQENARHSLGMSHOCILYNLAVNHALEEV 201
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 120 TLQCHHVHOGKFEVLGLSNVSWVEAEICTLCKKNGWIMPTVYOGMYNAITRQVETEL 179
QY 202 LPAQAYGLGVFAMSPILHGLLSGALKLAAGTAVKSAOGRAGVILPSLR-----PAIEAY 257
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 180 FPCLRHFGRLRYAFNPLAGLITGRK-----YQKDKGN-----PESRFGCPFSQLY 228
QY 258 -----EKFCRNIGEDPAEYGLAWLSRPGIAGAVIGPRTPEQLDSALK----- 300
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 229 MDRYWKKEHF-----NGIALV---EKALKTYGPTAPSMISAAYVWMYHHSQLKG 275
QY 301 --ASAMTLDEQALSELDEIFPAVASGGAAP-----EAW 331
DB 276 TQGDVAILGMSSTLEQLEONLALVEEGPLEPAVVADAFDOAW 315
```

```
RESULT 11
US-07-941-414-1
Sequence 1, Application US/07941414
Patent No. 5376544
GENERAL INFORMATION:
APPLICANT: LAZARUS, ROBERT A.
APPLICANT: HURLE, MARK
APPLICANT: ANDERSON, STEPHEN
APPLICANT: POWERS, DAVID B.
```



ADDRESSEE: HOWREY & SIMON  
STREET: 1299 PENNSYLVANIA AVENUE, N.W.  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: US  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/585,595  
FILING DATE: 16-JAN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/584,019  
FILING DATE: 11-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ADERBACH, JEFFREY I.  
REGISTRATION NUMBER: 32680  
REFERENCE/DOCKET NUMBER: 6137-0014 CIP  
TELEPHONE: (202) 383-7451  
TELEFAX: (202) 383-6610  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 278 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: 2,5 DKG REDUCTASE A  
INDIVIDUAL ISOLATE: CORYNEBACTERIUM SP.  
US-08-585-595-1

Query Match 10.2%; Score 177; DB 1; Length 278;  
Best Local Similarity 25.1%; Pred. No. 4.3e-11;  
Matches 76; Conservative 44; Mismatches 117; Indels 66; Gaps 11;  
QY 31 RVEDDDALRLMDHARDGNCINCLTDADMYGRLTKGHEELVGRRLAOGGREGRETVLATK 90  
DB 23 KVPADQRAVEALEVEGYRHIDTAIYG-----NEEGGAIAASGJARDDLFTTK 75  
QY 91 VGGEMSERVNDGSLARH-----ITASCEGSLRGLVDHIDYOMHIDRSAP-WDEVWQA 145  
DB 76 LM-----ND-----RHGDDEPAATAESLAKLADQVDLYVHPTPPAADNVYHAMEK 123  
QY 146 MDSLVAAGKVSYYGSSNFAGMHIAAOENAAARRHSLGMSHOCCLYNLAVRHALEVLPA 205  
DB 124 MIELRAAGLTRSIGVSN-----HLVPHLERIVA--ATGVVPAVNOIELHPAYQOREITDMA 177  
QY 206 QAVGLGFANSPHLGCL--LSGALEKLAAGTAVKSAOGRQVLLPSLRPAIEAYEFCRN 263  
DB 178 AAHVVKIESWGPILOGGYDYLFGAEPVTAATAA----- 209  
QY 264 LGEDPAEVLGLAVLRSRGICGAVIGPRT--PEQDLSALKASAMTLDQALSELDEIFPAV 321  
DB 210 HGKTPPAQAVLRMHLQK---GFVFFPKSVRRERLEENLDVDFDLDTEIAAIDAMPDGD 265  
QY 322 ASG 324  
DB 266 GSG 268

RESULT 14  
US-08-749-337-1  
Sequence 1, Application US/08749337  
Patent No. 5912161  
GENERAL INFORMATION:  
APPLICANT: LAZARUS, ROBERT A.

APPLICANT: HURLE, MARK  
APPLICANT: ANDERSON, STEPHEN  
APPLICANT: POWERS, DAVID B.  
TITLE OF INVENTION: ENZYMES FOR THE PRODUCTION OF  
TITLE OF INVENTION: 2-KETO-L-GULONIC ACID  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HOWREY & SIMON  
STREET: 1299 PENNSYLVANIA AVE., N.W.  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,337  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ADERBACH, JEFFREY I.  
REGISTRATION NUMBER: 32,680  
TELEPHONE: (202) 383-7451  
TELEFAX: (202) 383-6610  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 278 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-749-337-1

Query Match 10.2%; Score 177; DB 2; Length 278;  
Best Local Similarity 25.1%; Pred. No. 4.3e-11;  
Matches 76; Conservative 44; Mismatches 117; Indels 66; Gaps 11;  
QY 31 RVEDDDALRLMDHARDGNCINCLTDADMYGRLTKGHEELVGRRLAOGGREGRETVLATK 90  
DB 23 KVPADQRAVEALEVEGYRHIDTAIYG-----NEEGGAIAASGJARDDLFTTK 75  
QY 91 VGGEMSERVNDGSLARH-----ITASCEGSLRGLVDHIDYOMHIDRSAP-WDEVWQA 145  
DB 76 LM-----ND-----RHGDDEPAATAESLAKLADQVDLYVHPTPPAADNVYHAMEK 123  
QY 146 MDSLVAAGKVSYYGSSNFAGMHIAAOENAAARRHSLGMSHOCCLYNLAVRHALEVLPA 205  
DB 124 MIELRAAGLTRSIGVSN-----HLVPHLERIVA--ATGVVPAVNOIELHPAYQOREITDMA 177  
QY 206 QAVGLGFANSPHLGCL--LSGALEKLAAGTAVKSAOGRQVLLPSLRPAIEAYEFCRN 263  
DB 178 AAHVVKIESWGPILOGGYDYLFGAEPVTAATAA----- 209  
QY 264 LGEDPAEVLGLAVLRSRGICGAVIGPRT--PEQDLSALKASAMTLDQALSELDEIFPAV 321  
DB 210 HGKTPPAQAVLRMHLQK---GFVFFPKSVRRERLEENLDVDFDLDTEIAAIDAMPDGD 265  
QY 322 ASG 324  
DB 266 GSG 268

RESULT 15  
US-08-336-198C-3  
Sequence 3, Application US/08336198C  
Patent No. 586382  
GENERAL INFORMATION:  
APPLICANT: Hallborn, Johan  
APPLICANT: Penttila, Merja



APPLICANT: Ojamo, Heikki  
APPLICANT: Keranen, Sirkka  
APPLICANT: Hahn-Hagerdal, Barbel  
APPLICANT: Waldfidsson, Mats  
APPLICANT: Airaksinen, Ulla  
TITLE OF INVENTION: Xylose utilization by recombinant yeasts  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 N. Washington St.  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22046-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/336,198C  
FILING DATE: 03-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 318 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-336-198C-3

Query Match 9.3%; Score 161.5; DB 2; Length 318;  
Best Local Similarity 23.5%; Pred. No. 2.5e-09;  
Matches 70; Conservative 49; Mismatches 104; Indels 75; Gaps 12;  
QY 59 GRLYKG---HTELVG---RWLAOGGGRREDYVLTAKVGEMSERVNDGSLSRHII 110  
DB 38 GRLFDGAEEDYANERKLVAGVKKAIIDEGIVKREDLFITSKLMNNYHHPDN-----VE 89  
QY 111 ASCEGSLRLGVHDHIDVQMH-----HID-RSAPMDEVQQA 145  
DB 90 KALNRTLSLDLDVYDLPLIHPPVTFKFPVPLEEKYPPGFGCGKDNFDEVDVPILETWKA 149  
QY 146 MDSLVASKVSVYSSNFAGMIIAAOENAAARHSISGMVSHOCLYNLAVRHALE---VL 202  
DB 150 LEKLVKAKIRISIGVSNPGALLDLRLGATIKPSVLQVEH-----HPYLDQPRLI 200  
QY 203 PAQAVYGLGVFRWSPPLHGCLSGALEKLAGTAVKSAQGRAOVLPLSLPAIEAYEKFCR 262  
DB 201 EFAOSRGIAVPAYS-----SFGPOSFVELNQGRLNLTSPLEFN--ETIKATAA 246  
QY 263 NIGEDPAEYGLAWVLRPGIAGAVTGPRT---PEOLDASLKASAMTLDQALSELDEI 317  
DB 247 KHGKSPAOVLLRMSSOR---GIALIPKSNVYPRILENK-DVNSPDLDEQDFADIATKL 299

Search completed: June 22, 2001, 09:24:49  
Job time: 110 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2001, 10:02:37 ; Search time 1502.48 Seconds

(without alignments)  
10284.517 Million cell updates/sec

Title: US-09-463-705A-1\_COPY\_48\_1046

Perfect score: 999  
Sequence: 1 ctcgaaccgcgctccgcgcg.....gtccgcgcgtcgtgtgtcat 999

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_cm:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_htgc\_hum:\*  
20: em\_htgo\_inv:\*  
21: em\_htgo\_rod:\*  
22: em\_htg\_hum1:\*  
23: em\_htg\_hum2:\*  
24: em\_htg\_hum3:\*  
25: em\_htg\_hum4:\*  
26: em\_htg\_hum5:\*  
27: em\_htg\_hum6:\*  
28: em\_htg\_hum7:\*  
29: em\_htg\_hum8:\*  
30: em\_htg\_inv1:\*  
31: em\_htg\_inv2:\*  
32: em\_htg\_other:\*  
33: em\_htg\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_cm:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_vi:\*  
53: gb\_sts1:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_v11:\*  
59: gb\_v12:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_ro1:\*  
95: gb\_ro2:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length-DB	ID	Description
1	999	100.0	3439	3 SEY14332
2	999	100.0	3439	9 AX000482
3	999	100.0	3480	3 SE077454
4	999	100.0	3756	9 AR092721
5	673.4	67.4	47981	2 AF263245
6	593.6	59.4	5908	2 AF147704
7	511.6	51.2	9530	1 AB032523
8	486.2	48.7	47981	2 AF263245

Adapted

```

CDS      /gene="eryC"
          complement(<1. 13)
          /gene="eryC"
          /function="Involved in erythromycin A biosynthesis"
          /codon_start=1
          /transl_table=11
          /product="O-methyltransferase"
          /protein_id="CA474708.1"
          /db_xref="GI:4388593"
          /translation="MSVK"
          complement(45. 1046)
          /gene="eryBII"
          complement(45. 1046)
          /gene="eryBII"
          /function="Involved in mycarose biosynthesis"
          /note="putative"
          /codon_start=1
          /transl_table=11
          /product="grp-4-keto-L-6-deoxy-hexose 2,3-reductase"
          /protein_id="CA474709.1"
          /db_xref="GI:2765406"
          /db_xref="SPTREMBL:O33936"
          /translation="MTPTAAHYRLVGRSALLTSRLICTGVNFSGRVEDDQALRMDMDA
          RDGRIINCLDPTDMYGMRLYKCHTEBELVGRMLAOGSGGREDIVLTAKVGGSEKSEVYDSS
          GLSASHIILADESGSLRRLGVHDIDVYOMHHIDRSPWDEWQADSLVASGKSVYSSS
          NFGALWHLIAAONENARNRSHSGNVSHOCTNLAVRHALEVLPAQAQYGLGVFPMSPL
          HSGFSGMTEKLAICTAVKSAQGRQAVLLPSLRPAIEAYEKFCNLGEDPAEYVLIAMV
          LSRPIACAVIGRPPEQDSALRKASAMTIDQALSELDEIFPAVASGAAPEMWLQ"
          complement(1043. 2308)
          /gene="eryCIII"
          complement(1043. 2308)
          /gene="eryCIII"
          /function="Involved in desosamine biosynthesis"
          /note="putative"
          /codon_start=1
          /transl_table=11
          /product="glycosyltransferase"

```

[illegible]



[illegible]

RESULT	3
SEU77454/c	
LOCUS	3480 bp DNA BCT 10-NOV-1997
DEFINITION	Saccharopolyspora erythraea DEBS111 (ery111) and erythromycin O-methyltransferase (eryG) genes, partial cds, and erythromycin desosaminyltransferase (eryCII) and eryBII (eryBII) genes, complete cds.
ACCESSION	U77454
VERSION	U77454.1 GI:2599274

KEYWORDS	
SOURCE	Saccharopolyspora erythraea.
.ORGANISM	Saccharopolyspora erythraea Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Pseudonocardiaceae; Saccharopolyspora.
REFERENCE	1 (bases 1 to 3480)
AUTHORS	Hutchinson,C.R. and Katz,J.L., Sumners,R.G., Donadio,S., Staver,M.J., Wendt-Pienkowski,E.,
TITLE	Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea that are involved in L-mycarose and D-desosamine production
JOURNAL	Microbiology 143 (Pt 10), 3251-3262 (1997)
MEDLINE	98015410
REFERENCE	2 (bases 1 to 3480)
AUTHORS	Summers,R.G., Staver,M.J., Donadio,S., Wendt-Pienkowski,E., Hutchinson,C.R. and Katz,J.L.
TITLE	Direct Submission
JOURNAL	Submitted (05-NOV-1996) D-47N AP9A, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064, USA
FEATURES	Location/Qualifiers
SOURCE	1..3480
	/organism="Saccharopolyspora erythraea"
	/strain="NRRL 2338"
	/db_xref="taxon:1836"
gene	1..57
CDS	<1..57
	/gene="eryAIII"
	/codon_start=1
	/transl_table=1
	/product="DEBSIII"
	/protein_id="BAB84065.1"
	/db_xref="GI:2599275"
	/translation="HADAIARHIDAWLGGSN"
gene	43..1136
RBS	/gene="eryCII"
	43..47
	/gene="eryCII"
	/note="putative"
CDS	54..1136
	/gene="eryCII"
	/codon_start=1
	/transl_table=1
	/product="ErycII"
	/protein_id="BAB84066.1"
	/db_xref="GI:2599276"
	/translation="MTTDDRAGLROLOMIRGLHMWGSNGDPYPMILCGHDDDDPORRRYRSREGSVRRRTETWVAADHATARAVIDDPAFGRGTPENMRKAPGAEMAOPE RDVHAASMEGEVPDYGELAESFAGLLGACARDLDVDVFAMQVVOGMTAVLAGAVLRGAAMDARVSIDDQAOLSPQQLAVTEAAVAALPADRALPARGAEANTVVDAVLAVSAEPGLARIADDPAQAORTAEVLRHLPALHLEERTATAEVRGAEHVIGEEEVVVVVAANARDEVEFAEPDRILDVREDPDADRALSARHGHPGLEELVTALATRAAKAAPALPG LTSGPVARRRRSRPVLTGNRCPEVL"
gene	1135..2412
RBS	/gene="eryCIII"
	1135..1138
	/gene="eryCIII"
	/note="putative"
	1147..2412
	/gene="eryCIII"
	/codon_start=1
	/transl_table=1
	/product="desosaminyltransferase"
	/protein_id="BAB84067.1"
	/db_xref="GI:2599277"
	/translation="MRVFSMSASKSHLGLPLPMAAFRAAGHEVRVVASPALTEDITIAAGTAIVGVGDVLDVFMFHAGDIIDVYSLSDFSDPATLFMHNLGMOTVLPPTFYALMSDPILEGVNSCRKRKPDLVIETLEPAATVGTPHARHLGGDTTFTRKQTLGLPDQPEHRERGLAEMLTTLEKTGCPADEEYVGOWITDPAARAIRLETGLNFTVMRYVDVNGPSVVPDWMLHDEBERRRVCLLTGISSEKENSIGOVSIPELLGAVDVADELATEDDAQOLEGVANTPHNVRTVGVPMHALPTCAATVHHGGPGSMHTAAIHLGVPOVIIILTDGTVARQTOEFEGAGIALPVELTLPOLRESVKRYLVDDPHRGAAR

```

gene      2394. 3410. MRDDMLAEPSPAEVVGICEELAAGRRE"
/gene="eryB1"
RBS      2394. 2399
/gene="eryB1"
CDS      2409. 3410
/note="putative"
/gene="eryB1"
/codon_start=1
/transl_table=1
/product="EryB1"
/protein_id="AAB84068.1"
/db_xref="GI:2599278"
/translation="MTTDAATVRLGSGSALLTSRLMLGTVNFSGREDDALRLMDIAR
RDRIINCLDITADMTGMRLYKGTIELVGRMTAOGGREDIVLTAKVGGESEEVNOS
GLSRHIIASCEGSLRLRGYDIDVYDMHILDRSAPDEWQADSLVASGKSVYSS
SNFGNIAAQAENARRSHSLGNSHOCYLILARNAIEVLPAQAQAGLGFAPMSPI
LSRGISGLEKTLAAGTAVKSAQGRAOYLIPSLRAPIEAIEYERCFNLSGEDPAEVLAW
LSSRGINGAVTIGPRTPEQDLSALKAAMTLDQALSELDELTFPAVASGARPEMLQ"
4429. 3480
/gene="eryC"
RBS      3429. 3433
/gene="eryC"
CDS      3442. >3480
/note="putative"
/gene="eryC"
/codon_start=1
/transl_table=1
/product="erythromycin O-methyltransferase"
/protein_id="AAB84069.1"
/db_xref="GI:2599279"
/translation="MSVKQKSAIOLDLV"

```

[illegible][illegible][illegible]







QY 430 gtacaggactggtgagagaccatgcccagaggtgagcgagcgagctctcctgagc 489  
|||||  
Db 8645 gTACAGGACATGCTGGGAGATGATGCCGAGACAGTTCGGCGTGCCTCTCTCGGC 8704  
QY 490 gggcgagatgtgcagcccgagagttcgacgagccgagactagagagacttgcgcgtgc 549  
|||||  
Db 8705 GGGCGGACATGTGCCAGCCCGCAGGTTGAGAGACGACATACCCTCCACTGCC 8764  
QY 550 gaagagcgttccatgtgagccgacacactctctcccaagcgagccagcgtgcatgtgtg 609  
|||||  
Db 8765 GACCAAGATGTTCCGGCGGCTGCCACACCTCTGCCACAGGTGCGCGCGTGCATGTGTG 8824  
QY 610 catctggttagacgttcgattgtgtgcagccagccagcctgcgacgatactctcgagagagc 669  
|||||  
Db 8825 CGCTGTGATGATGTGATGTGTGATGATGCCCGGAGCGGAGGAGTTCGCGAGCGCGC 8884  
QY 670 gatgatgtgcgagccagccagccagcctgtctgttcagcagcgtctcactcgcgcgcgac 729  
|||||  
Db 8885 GACGATGTGTCGGGCGAGAGACCCCGCGTGTGATCCGCTCTCATCTGCTGCTCCAC 8944  
QY 730 ctgtgtcgcagcagaggtgtctcctgcgcgcgtccgcgcctctgagccagcactgtccac 789  
|||||  
Db 8945 CTGTGTCGACAGAGACGCTCTCTGCTGCTGACACCTCCGCGCGGCGAACCACGCTCCGAC 9004  
QY 790 cagctcctcggtgtgtgcctctgttagagcgccagccagcgtacatgtcgcggtgttcgagga 849  
|||||  
Db 9005 GAGTTCCTCGTGTGCTGTGCTGTGATGACCCCGCCTACATGTGCGCGGTGTGATGAC 9064  
QY 850 gttaatgcgcgcggttccgcgcggtgtctcatcagagcgagcgagcgtctcctcgcgcgcg 909  
|||||  
Db 9065 GTTGACGCCCGCTCGAGGGCGTGTCTCATGACGCCGACGCGCTGCTGCTGCTGCTGCTG 9124  
QY 910 tccgctgaagtctacacgttgcgcagcagagcctgtgtgtgtagagcgagcgagcccgag 969  
|||||  
Db 9125 TCCACTGAGATTACGGGTGCCGAGCCAGAGTCGGCTGTGTGCAACGCCGATGCTCCGAC 9184  
QY 970 ccgacggtcgtgcgcgcgtcgcgtcgtgtgt 996  
|||||  
Db 9185 GCGTACCCGGGCGAGCCCGCGCTGCT 9211

RESULT 6  
LOCUS AF147704 5908 bp DNA BCT 08-FEB-2000  
DEFINITION Streptomyces fradiae NDP-hexose 2,3-enoyl reductase TylicII  
(tylicII), NDP-hexose 4-ketoreductase TylicIV (tylicIV), NDP-hexose  
3-C-methyltransferase TylicIII (tylicIII), mycarosyl transferase  
TylicV (tylicV), and NDP-hexose 3,5- (or5-) epimerase TylicVII  
(tylicVII) genes, complete cds.  
ACCESSION AF147704  
VERSION AF147704.1 GI:5305790  
KEYWORDS Streptomyces fradiae.  
SOURCE Streptomyces fradiae.  
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 5908)  
AUTHORS Bate,N., Butler,A.R., Smith,I.P. and Cundliffe,E.  
TITLE The mycarose-biosynthetic genes of Streptomyces fradiae, producer  
of tylosin  
JOURNAL Microbiology 146 (Pt 1), 139-146 (2000)  
MEDLINE 20121747  
PUBMED 10658660  
REFERENCE 2 (bases 1 to 5908)  
AUTHORS Bate,N. and Cundliffe,E.  
TITLE Direct Submission  
JOURNAL Submitted (30-APR-1999) Biochemistry, University Of Leicester,  
University Of Leicester, Leicester LE1 7RH, UK  
FEATURES  
source  
1..5908  
/organism="Streptomyces fradiae"  
/strain="T59235"  
/db\_xref="taxon:1906"

gene complement(80..1069)  
/gene="tylicII"  
/note="orf10\*"  
CDS complement(80..1069)  
/gene="tylicII"  
/function="biosynthesis of NDP-mycarose during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="NDP-hexose 2,3-enoyl reductase TylicII"  
/protein\_id="AAD41821.1"  
/db\_xref="GI:5305791"  
/translation="MSGMTVOLGKATVLSRLMCTVNSGHPVDDQAIRLMDALDR  
GVNCDVADIVGMRLYKHTBELVGMKLSGRDDVYVATRVGPMSDRVNDRGLSA  
RHVINSCEASLRLGLVDHIDLYOMHMDRTVMQMDQVLSGKRYTIGSSNFA  
GWHLAAGSEARBSGLGVSBQCYNLAVRAELEVLPAARAYGIVGVAWSPLGGL  
LSGALRKLAEGTAAVKSOGGROARTI.PALBDTIAREYERCAVRGRDPAEYGLMILSRP  
GVSAGVIGRTTGHVLSALRAYELESEHEHLEALFPVSGGEVPAWON"  
gene complement(1066..2109)  
/gene="tylicIV"  
/note="orf9\*"  
CDS complement(1066..2109)  
/gene="tylicIV"  
/function="biosynthesis of NDP-mycarose during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="NDP-hexose 4-ketoreductase TylicIV"  
/protein\_id="AAD41822.1"  
/db\_xref="GI:5305792"  
/translation="MSMPLVPGPARDRPLVVLGASGFLGSAAVAELIARRPRLRL  
VARBPSPVAGAVAEIEVRRDLSLPEVTEAVEDADAVHLVAHTGKRSVRAAGER  
SEHINVLMELEALRLPTPADATAPAVVLGSLTQAOEMEARHTEGTYAAAKLAER  
VLRADAEAGVAVGLRLFTYIGRSLTSGPGRGVAVYAGALAGDPTTMHDSVE  
RLDLYRDATATFTALHADOLRGKHWVYGRHRLDRVGTVAALAAEHTRGPV  
PVYVDPPEYAEVCDPRTDSDPSANRAVYGNRPRAEDGIAAIAVAAGADSEPE  
EGAGKRGGTERK"  
gene complement(2106..3362)  
/gene="tylicIII"  
/note="orf8\*"  
CDS complement(2106..3362)  
/gene="tylicIII"  
/function="biosynthesis of NDP-mycarose during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="NDP-hexose 3-C-methyltransferase TylicIII"  
/protein\_id="AAD41823.1"  
/db\_xref="GI:5305793"  
/translation="MPAVPREDOMTISACRVCGNRELLPVLGEGALGVPPRREE  
TVPISPLELWKSPOGGLVQLRHTPDPLMGEGYSGIRPEMIGHLRRKVAIR  
ELVDLGGDDLVLDIGSDATLLKAYEGRVLGIDPSGCFRELYPPHAILAEYFSR  
DVTAFRGRRARVITSTIMFYDLPPLAFMPDHYDVLADDCITWEOGYLRLAEAD  
AIDIVCHEHLEYALDOITEMABRACITVLRLELDYVGSGLCYTLARSSPHRDEA  
GPARIRARETEAKINTMAPFEFARVREVRDLRDLRSRAGRLTLGYASTKGN  
VILQYCGIERDLPICGEVSPKAGFTGTGTGIPVSEEDAKAMRPDOLLVPIWYRE  
GEVERERDPLACGGRLVPLPLRDYV"  
gene 3575..4960  
/gene="tylicV"  
/note="orf7\*"  
CDS 3575..4960  
/gene="tylicV"  
/function="catalyses the incorporation of mycarose during  
tylosin production"  
/note="glycosyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="mycarosyl transferase TylicV"  
/protein\_id="AAD41824.1"  
/db\_xref="GI:5305794"  
/translation="NAGLRPAGAGVPGCTPMPISPGKCVSAISRRLPYAPGPTPL  
GRRRAPDSCRMHAHTAFILIPAGHVPPTLGVAEELARHRTYALPEMDADRAV  
RGARAVTYPLDRERADMPKESDEYTDGEFLKYLEWLLDTTADTLPLESAFA

BASE COUNT

ORIGIN

853 a 2123 c 2078 g 854 t

gene

CDS

EDRRADVAAANDPESTFWTGLLLADKWDIVIRISTPYSASNEHNAHLHPPEEGAAOYDPA  
IELRARAELKLKEHGTSTDPAPAAATVOSGGGLFVPRYQYAGETEDRRHFFVGPAA  
PRASEHGMOREPDEGRLPYVWSLCTITNERGPIRACVAPRDRPAMILLVGLGSLAA  
GDCLPPEHNVLYVRDVPFLGDVLPHTDILLNHGCTSTAMEALAHVPIYVAMPEKPRPA  
TARRIAIEDLDGDMLLPGEVTAKEKLSIAQRVLVDIDRRKELRRMGEIRAGGAVAA  
DVIEGLLSPAA"  
5004. .5597  
/gene="tylcvii"  
/note="orf6"  
5004. 5597  
/gene="tylcvii"  
/function="biosynthesis of NDP-mycarose during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="NDP-hexose 3,5-(orf5)-epimerase Tylcvii"  
/protein\_id="AADA1825.1"  
/db\_xref="GI:5305795"  
/translation="MITEETVRDVARITPEEIPDARGSLYESLRVETLRNRRATHAIE  
IRQNVHVNRRNTLRGTHGHTVPPGCGKITVCVGAARTNVDLRVSGSPFEGHDVVG  
ODASGVAAYLIPDGLIGGYALADDTCCMNTLYTEGRTYVGMIIIDALDIPGLGLPWNLT  
EPPVRSQDAAPSLAAEAAAGTIPGEGCCLRAPM"

[illegible]

OY	662	caggagggatgatgtgtgcgagcccgaccccgcgttgttgtagcgcgtcatctcgcg	721
OY	722	ccgcgaaccttggctgcacagacgltgtccttcgcgcgcgttcgcgcgccttgagccagcac	781
Db	804	TGCGCGACCTTGGTGCGCAGCACACACGTTCGTCCGCCGCTCCGCCTGCC---GCCACAGCCAC	860
OY	782	ctgccaccagaactcctcgtgtgtgacctttgttagagcgcagccagtcgttacattcggcgtg	841
Db	861	GCTCCAGACAGATTTCTCCGTGTGCCCCCTTGTGAAGAACCCACAGCCCTAATATGTCCGCCCTG	920
OY	842	tcaaagcagattgatctccgcgcgtlcccgggccggtgtgtcattcaaaagcagcagcgttcgtcc	901
Db	921	TTCAACGCGATTGACCCCGGCTCAGAGCCCTTCGTTCATGAAAGCGATATCGCTGTCTGCG	980
OY	902	tcaaagcagcttcagcttgaagtltcaacgctgtccgagacagacgtctgtgttagcagcgcgaa	961
Db	981	GGGACTCTCCCGCTGAATGTAACCGTGCACCAGAGCGCGGTGACAGTAGTGTGCGCGCG	1040
OY	962	cgccccgagcgcagctgtcgtcgcgcg	985
Db	1041	CTCTCCAGTTCTCAGCATTCCTCG	1064
RESULT	7		
AB032523			
LOCUS	AB032523	9530 bp	DNA
DEFINITION	Streptomyces avermitilis avermectin biosynthetic gene cluster (orfI, aveBII, aveBIII, aveBIV, aveBV, aveBVI, aveBVII, aveBVIII), complete cds.	BCT	23-SEP-1999
VERSION	AB032523		
KEYWORDS	AB032523.1 GI:5921154		
ACCESSION	dntp-4,-keio-6,-deoxy-L,-hexose 2,3-reductase; dntp-6,-deoxy-L,-hexose		
3-O-methyltransferase; dntp-4,-keio-6,-deoxy-L,-hexose2;			
3-dehydratase; dntp-4,-keio-6,-deoxy-L,-hexose 3,5-epimerase;			
dntp-4,-keio-6,-deoxy-L,-hexose 4-reductase; glucose-1-phosphate			
thymidyltransferase; dntp-glucose 4,6-dehydratase; glycosyl			
transferase; reductase.			
Streptomyces avermitilis DNA.			
Streptomyces avermitilis			
Streptomyces avermitilis			
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
REFERENCE	1 (bases 1 to 9530)		
AUTHORS	Ikedu,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.		
TITLE	Organization of the biosynthetic gene cluster for the polyketide		
JOURNAL	antibiotic macroide avermectin in Streptomyces avermitilis		
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)		
MEDLINE	99380548		
2 (bases 1 to 9530)			
Ikedu,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.			
Direct Submission			
Submitted (17-SEP-1999) to the DDBJ/EMBL/Genbank databases, Haruo			
Ikedu, School of Pharmaceutical Sciences, Kitasato University,			
Microbial Chemistry, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641,			
Japan (E-mail:ikedu@mc.pharm.kitasato-u.ac.jp). Tel.:+81-3-5791-6242,			
Fax:+81-3-3444-6197)			
Location/Qualifiers			
1. 9530			
/organism="Streptomyces avermitilis"			
/db_xref="taxon:33903"			
160..873			
/gene="orf-1"			
160..873			
/gene="orf-1"			
/note="contains NAD(P)H binding domain; not function in			
avermectin biosynthesis"			
/codon_start=1			
/transl_table=1			
/product="reductase"			
/protein_id="BA04591.1"			
/db_xref="GI:5921155"			
FEATURES			
source			
gene			
CDS			

gene  
CDS  
/translation="MCVDRDGRGAGTADLARTGRARAAMEVADVSDGAMEREAEER  
VAETVGVVDLLVNNAGIMAGRLDTSVEDWQRTGLVNGVTHGCRLLGROMAERGO  
GGHIVTASAAAFQPTPAVPAATSKAAVLMLSECRARAEFGVGVSVCCGFVTS  
FASAMFAGVPRLEDERLRALFAGRCSSAEKVAVALRSVARDASAVVTATEARLSRL  
MSRFPRLRAAARMPDPS"  
931..2169  
/gene="aveb1"  
/protein\_id="BA84592.1"  
/note="transfer of oleandrose onto polyketide-derived  
acylcons"  
/codon\_start=1  
/transl\_table=11  
/product="glycosyl transferase"  
/protein\_id="BA84592.1"  
/db\_xref="GI:5921156"  
/translation="MSDIFLMSAPFPGWHPFSLAVAEELVHGHVTEVGAEMADA  
VRSVGDPLRYESAPFMDYMLTEAPNALDPTLXDCGMSKLRSEEHCKGVVDL  
VADTATSLNCGRVLAASMSRAMTYIPFASNGRSTQSVLDPDSQVSAFPPRS  
EOMELFGALVPRLAELLVSKITEVPDPLSGPDMVCLPRAFOYAGDTEDERF  
AEVGPCLGRKRGENTPPGSGHPVLISLGTVENROLSEFFETVRAFTDVPHVVIS  
LGKGVDPVLRPLRPVNEVRWPHAVLEHARALVTHGCTGSVMELAGCCVLYMP  
LSRDAQVGRTRIAELGIRMGVQPEEVATTLRRHVLDLISDAITRQVQMRATVEA  
GGAIRADETERFLRTRRH"  
complement(2075..3142)  
/gene="aveb11"  
complement(2075..3142)  
/note="function in oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="dtdp-glucose 4,6-dehydratase"  
/protein\_id="BA84593.1"  
/db\_xref="GI:5921157"  
/translation="MTGAGRTGSHFVRRLTGAVPAFTGAEVYVLDKITYAGRLNT  
AVTGPSLITFVHGDCDGPLVADLMDGSDMYVHFAESHVDRSVADAEPRTNVLG  
TITLRATDAADRYTISTDEVTGISTGSTEAPLEPNSPYSASASDLKARS  
FRTTGLPVIITRCSNNYGHQPEKLIPIRFVTHLNGTKVPLYGGEVNRMLHARD  
HCRGIALVABERGPETVHIGGTLSNELRALDLGVMSWEPEYTDKRGHNR  
YSLDISKISAELEGYARVPPEEGLAQTQVQVYENRTLMPELTARPELPVSDGSAET  
ASRRLPACGRPRPMPASA"  
complement(3139..4038)  
/gene="aveb111"  
complement(3139..4038)  
/note="function in oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="glucose-1-phosphate thymidyltransferase"  
/protein\_id="BA84594.1"  
/db\_xref="GI:5921158"  
/translation="MKGIYLAGTGSRLYPLTRALSQKLLPYVDKPMIYYPYSLMIG  
GIKDLIVISSPHLEQFRLLDGSRLGINTVIAAQORGGIAEFLIGADFGGOOV  
SLVGNIFHGMGFSHLRSHTRDYGCVLFGVAVDPERGCVGASGKLLSVK  
PPAPRSNTAITGLYXNDVYITVARGICRSARGELITPYNRAYLAEGARALVDLGRG  
ETWLDAGTHDSLHMAQIYQVLEKROGVRIACLEIAFRMLGLIDADCYLRGVELAGS  
4164..5195  
/gene="aveb12"  
/protein\_id="BA84595.1"  
/db\_xref="GI:5921159"  
/translation="MGRSVCPPRPPIGLKSMITGMDPLVYVLGASGYIGSAVAA  
ETLRWVLLRLVARRPVPGGAETETRTADLTAASEVALAVDDADVILHVARLT  
OGAAMAAESDPAERVNVGVMDVVAALRSGRAGRPVYVAGSVYQVGRGRVGD  
SEDEVTAYAROKLDAERTLSATVEGALRSGLSLPLVYAGPGPGNGVQVQAVL  
RLADEALVWNGSVYERDLVHVEDVAAQFVSGIALADLGRHMLIGSGRPVPHL  
PGAIAGVSAKTRGRAVPTAVDPAMATADPHGIVYDSSAFRAVTKMRPLSLDQES

gene  
CDS  
/translation="LDMHVAAYV"  
complement(5249..5926)  
/gene="aveb1"  
complement(5249..5926)  
/note="function in oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="dtdp-4-keto-6-deoxyhexose 3,5-epimerase"  
/protein\_id="BA84596.1"  
/db\_xref="GI:5921160"  
/translation="MTLGRPRSSADRPAPGARTAGVVRRLVENGAEFPPTV  
PFDERLEVTYQDEPLSEAVGHREPTAQCSVSRGVKGVHFTPATPGQAKVYHC  
ARGRALDVFVLDRTGSPFGQMDSVLLDERFRSVYLPVGVGHAFVALLDDQAMYL  
SSGVYQNHALLSPEDDPLPLGHHLGRAPILSERDRAPTLQOALRGMLPEYRAS  
RGIDKEL"  
complement(5988..7397)  
/gene="aveb11"  
complement(5988..7397)  
/note="function in oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="dtdp-4-keto-6-deoxy-L-hexose 2,3-dehydratase"  
/protein\_id="BA84597.1"  
/db\_xref="GI:5921161"  
/translation="MSVRADADHTERTAHRAARRPARVPRPLRRGRHRRRTSLDA  
FTGMMTRBSGAHREPRYERIPFGMDMSFPGCTGNLAHSGRFSVEGCHLVGDCOP  
PERQDPIHQPEIGITLAKKFDGYLHFLMAKKEPBNINLVQUSPTVQARNSNTK  
VHGGAUVKYLEYETOPRRATVVVDVLOSQHGAMFHKRFRNRTVETDEVDPLDDEFRW  
LTLGOIGELMNDNLVMDARTVLACLPPEDEPAHMRSDADELSYVAERSRSHVA  
RRVPLAGIPGWTGAEISIAHADRFRVAVAEASNSRDAAMTOPLIPCCGHSTAF  
LRRIGGVNHLAHGVEGGLDTEIGTQVTPYPRVNHLPGRAPRLDLLEAP  
DLIRRAVHSEEGSRHLNARVYFPEADESAPNDPPRGYKWCITPQTLQLRGYR  
VWQARTLSLITTRAVEL"  
complement(7394..8167)  
/gene="aveb111"  
complement(7394..8167)  
/note="function in oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="dtdp-6-deoxy-L-hexose 3-O-methyltransferase"  
/protein\_id="BA84598.1"  
/db\_xref="GI:5921162"  
/translation="MTREFTDPRIVHESQOBARAEQVTLVDSPIPPYLLDNL  
SVYMRNOLADLSDALRYMLPEVDGPMGEGVLGRHLATLTLRSIYEPYNSLR  
VIGFTFTGFPDIDEDVSTSAVGRFVAPGGEVHNLQVLAANEANPYGHTQSRF  
VVOGDVRETVPOYLAHPHTVIALAFDIDLTPRELVDVTPHLLRTRESILAFELT  
HRKMPGERTALSEAFGLDHAPLROLPGRPPVIYMTWA"  
complement(8254..9297)  
/gene="aveb1111"  
complement(8254..9297)  
/note="function in oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="dtdp-4-keto-6-deoxy-L-hexose 2,3-reductase"  
/protein\_id="BA84599.1"  
/db\_xref="GI:5921163"  
/translation="MPTTAEPAPVQSDNSNAPLTELGRRLAISRLATGVNIG  
RVEEPARRLMHOLAQGITLPTANTYGMVHKGTTEVITGRMLADRPAREQVYLA  
TRVGDMGSPNDHGLSVRNIVAACDASLRRLTMDIDYOLHNDIRAGMDEVOQAM  
DLITOGKRVYVSSNFGAMDJASAOEARRRNALGLASEOCVUNLVTNHALELVPA  
ASAYGVLYWSPGLGGLGVLRTRRENTAVKSAQGRAVELHNRRTITAYVEDCA  
DHGIDPANHGMAYLSRPGVTLYVIGPRIEQVNDGALNHLRTPLRPPVLAIRBELFPP  
VGRGSAAPAMUS"  
BASE COUNT 1341 a 3029 c 3601 g 1559 t  
ORIGIN  
Query Match 51.2%; Score 511.6; DB 1; Length 9530;  
Best Local Similarity 70.6%; Pred. No. 1.6e-46;



[illegible]

Db	Accession	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE ORGANISM	REFERENCE
Db	46408	CAATGACACCTTCGAGGGGGTGTCTGGCCAGAGAGTCTTGACGAACACCTGTACGCTTCAC	SCD19/c	22991 bp	DNA	Streptomyces coelicolor A3(2)	Redenbach, M., Kleser, H.M., Denapalte, D., Eichner, A., Cullum, J.,
Qy	241	cgccgagcgaggaaggaagcaaacagcaactctgcgaagcgccctgcgcgaacttaaccgagct	ABC transport system ATP-binding protein; ABC transport system integral membrane protein; aminotransferase; DNA invertase; integral membrane protein; integrase; nucleosidyltransferase; oxidoreductase; oxidoreductase lipase; tetr family transcriptional regulatory protein; two component system response regulator; two component system sensor kinase.	AL392149.1	GI:10129746	Streptomyces coelicolor A3(2).	1 (bases 1 to 22991)
Db	46348	CGCCGACCGCAGTGTCCCTCTGGCTGTCCACAGGCCCTGTGGAGTGGGGCCCGACCG--	Streptomyces coelicolor A3(2)			Bacteria: Firmicutes: Actinobacteria: Actinobacteridae: Streptomycetaceae: Streptomycetaceae: Streptomyces	
Qy	301	gcccggagcagcttccagcgctccgctctgagcaagcgccgctgtcagcgaggaacagc					
Db	46290	-----TCGCCGCCAGACAGACCCTGGAGAGCGGCTTGCGCTTGGC					
Qy	361	gaagacgcagaccgctagagcctctgcagcgagcgagcaacacttccagctcgcgctgcagac					
Db	46252	GAAGACCCCGACCGCCGAGACCTGTGCGCGGGGAGAGACTCTCAGATTCGGATATGGCGCA					
Qy	421	cgccagctgttacaagcacttgtgtggagacacatgcccaaggagatgtcgagcgagcggtt					
Db	46192	CGTACAGTCTTACCGACACGTGGGGAGACACAGGCCGAGAGCGGTGACGCGGACGGCTGTG					
Qy	481	ctctctgcgagcgagcgagatgtgcacgcgcgcgaatttcagcagcgccgaacttagaagactt					
Db	46132	CTCTTGGGCGCGACAGTGTGCTTCCTCCGGGAGAGCCCGACGACCCGAGTACACACTT					
Qy	541	gcgcgtgtgcagagcgctgtccatagcctgtccacacactgttccacagcgcgagacgagct					
Db	46072	TCCGGCGGCCACGAGGGGCGTCCACGCCGTCGACAGACTGTGTCCAGCGGTTCACCCGGTTC					
Qy	601	gattgtgtcatcttgtttagacgttcgatgtgtgcagaccccaagcctgcgacgataccctc					
Db	46012	CACCCGGGGGAGGTGAAGGACGTGCAGCTGTGCACACCGAGACGCCGACGGAGACCTTC					
Qy	661	gcagagagcgataitgtgcgcgcgcgaagccgcgcgtgtgttagagcgctgcgtcatctc					
Db	45952	ACAGAGGCGCATATCTGCGCGGGGAGAGAGCGGCCGCCGACCTTCCCGCCGGTGG					
Qy	721	gcgcgcgaacttgctgcagacaggtgtgtccctgcgcgcgtccgcgcgcctgagcagca					
Db	45892	GACACCCACCGTACCGAGACAGAGAGGAGTCTCTCCCGCAGCGGCTGTCCCGGACCAACA					
Qy	781	cctgcaccacagctcctcgtgtgtgcctttagagcgcgacccgatactatgtcgcgt					
Db	45832	CCGCGCGACGACTCTCCTGCGCGACCTGGGCGACTGTGCGACGAGAGTGTGTGTGGCGGT					
Qy	841	gtcgaagcagttatgtcgcgcgtgccggcgctgtgttcatacgaagcgacgcgttcgtc					
Db	45772	GTCGAGGCACTTACACGCCCGCGAGAGGGCGTGTGCAGACGACTGCTCGTCCGCG--					
Qy	901	ctcgaagcgctgcgcgttaagttacacgttcgcgcgcgcgaagacccctgcgtgtgagcaagcgga					
Db	45714	-TCCTCTGTCGCGCGCGAGGGCTGTCCCTTACCCAGACCGACCGCTGTGTAGCAGGGCGCA					
Qy	961	acgcgcgaagcgacgtgcgtgcgcgcgcgttcgcgtgtgc					
Db	45655	CCGGCGAGCCGGACGATGGGTGGCTGTGGTGCATTC					

TITLE	Kinashi,H. and Hopwood,D.A.
JOURNAL	A set of ordered cosmids and a detailed genetic and physical map
MEDLINE	for the 8 Mb Streptomyces coelicolor A3(2) chromosome
REFERENCE	Mol. Microbiol. 21 (1), 77-96 (1996)
AUTHORS	2. (bases 1 to 22991)
JOURNAL	Seeger,K.J. and Harris,D.
AUTHORS	Unpublished
JOURNAL	3. (bases 1 to 22991)
AUTHORS	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
JOURNAL	Direct Submission
AUTHORS	Submitted (11-SEP-2000) Streptomyces coelicolor sequencing project,
JOURNAL	Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
AUTHORS	CB10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
JOURNAL	David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
AUTHORS	Colney, Norwich, Norfolk NR4 7UH, UK
JOURNAL	Notes:
AUTHORS	Streptomyces coelicolor sequencing at The Sanger Centre is funded
JOURNAL	by the BBSRC and Beowulf Genomics
AUTHORS	Details of S. coelicolor sequencing at the Sanger Centre are
JOURNAL	available on the World Wide Web.
AUTHORS	(URL: <a href="http://www.sanger.ac.uk/Projects/S-coelicolor/">http://www.sanger.ac.uk/Projects/S-coelicolor/</a> ) CDS are
JOURNAL	numbered using the following system eg SC7B7.01c. SC (S.
AUTHORS	coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
JOURNAL	strand).
AUTHORS	The more significant matches with motifs in the PROSITE database
JOURNAL	are also included but some of these may be fortuitous. The length
AUTHORS	in codons is given for each CDS.
JOURNAL	Usually the highest scoring match found by fasta -o is given for
AUTHORS	CDS which show significant similarity to other CDS in the database.
JOURNAL	The position of possible ribosome binding site sequences are given
AUTHORS	where these have been used to deduce the initiation codon. Gene
JOURNAL	prediction is based on positional base preference in codons using a
AUTHORS	specially developed Hidden Markov Model (Krogh et al., Nucleic
JOURNAL	Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
AUTHORS	of Bibb et al., Gene 30:157-66(1984) as implemented at
JOURNAL	<a href="http://www.nih.go.jp/jun/cgi-bin/frameplot.pl">http://www.nih.go.jp/jun/cgi-bin/frameplot.pl</a> . CAUTION: We may not have predicted the
AUTHORS	correct initiation codon. Where possible we choose an initiation
JOURNAL	codon (atg, gtg, ttg or (atc)) which is preceded by an upstream
AUTHORS	ribosome binding site sequence (optimally 5-13bp before the
JOURNAL	initiation codon). If this cannot be identified we choose the most
AUTHORS	upstream initiation codon.
JOURNAL	IMPORTANT: This sequence MAY NOT be the entire insert of the
AUTHORS	sequenced clone. It may be shorter because we only sequence
JOURNAL	overlapping sections once, or longer, because we arrange for a
AUTHORS	small overlap between neighbouring submissions. Cosmid D19.
JOURNAL	Location/Qualifiers
AUTHORS	1..22991
JOURNAL	/organism="Streptomyces coelicolor A3(2)"
AUTHORS	/strain="A3(2)"
JOURNAL	/db_xref="taxon:100226"
AUTHORS	/clone="cosmid d19"
JOURNAL	complement(1..358)
AUTHORS	/gene="SCD19.01c"
JOURNAL	complement(<1..358)
AUTHORS	/gene="SCD19.01c"
JOURNAL	/note="SCD19.01c, unknown(fragment), len: >119 aa.
AUTHORS	Contains two TTA leucine codons, possible targets for b1da
JOURNAL	regulation"
AUTHORS	/codon_start=1
JOURNAL	/trans_table=11
AUTHORS	/product="hypothetical protein SCD19.01c (fragment)"
JOURNAL	/protein_id="CAC08302.1"
AUTHORS	/db_xref="GI:10129747"
JOURNAL	/translation="MORTYGLKTELVDRLNGPLEGVTDLEAQAQLTVLHTELENGST
AUTHORS	GGGEKSLNEMEAALRTLKPLLLRLKLDLKAAPDFQSFRRKYWIREMGGGGGYAKRRR
JOURNAL	SYDGLFLEYPVERKIDEM"
AUTHORS	1..98
JOURNAL	/note="nominal overlap with Streptomyces coelicolor cosmid
AUTHORS	SCD12A"
JOURNAL	complement(35..37)
AUTHORS	/gene="none"
JOURNAL	gene

```

misc_feature complement(35..37)
/gene="none"
/note="TTA leucine codon. Possible target for bldA regulation."
label=TTA
138..143
/note="Drai site matched TTTAA"
/misc_feature complement(140..142)
label=Drai
/misc_feature complement(140..142)
/note="TTA leucine codon. Possible target for bldA regulation"
label=TTA
gene complement(140..142)
/gene="none"
gene complement(603..1031)
/gene="SCD19.02c"
CDS complement(603..1031)
/gene="SCD19.02c"
/note="SCD19.02c, unknown, len: 142 aa. High content in alanine, arginine and leucine amino acid residues"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCD19_02c"
/protein_id="CAC08303.1"
/db_xref="GI:10129748"
/transl_table="MTNNHQAQAEFTRIRSAMORLLAGSVPDGLKGVKSICTLIAGVPR
ATLYTTPYHLKAEEPORTAODAGOODPTLADIERLKEAVATLRERLSRKNVELDA
LKEPFWALSRIAAQOADIITSLRRELEAQRRLRPVRR"
complement(1028..3199)
gene "SCD19.03c"
complement(1028..3199)
/gene="SCD19.03c"
/note="SCD19.03c, hypothetical protein, len: 723 aa; similar to some plasmid derived hypothetical proteins, e.g. TR-085873 (EMBL:AF079317) Splingomonas armaticivorans hypothetical 81.3 kDa protein (plasmid pNL1), 741 aa; fasta scores: opt: 421 z-score: 478.1 E(): 4e-19; 29.7% identity in 750 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC08304.1"
/db_xref="GI:10129749"
/transl_table="MTTHASSPALAGRPHGIVLETAGLRREGSPRVDDVMY
DLTGADAPVYMSTHRKLIDLTATIIINPMROYAEYIMARLAPIHPDYAVPOAFRP
LAPNSWSELNHLAMEHLTAVGITISVSQVRHNCHQAYILAAASATDDRLISPRT
TVAMRIHQPIELVAEILSDCYORGFPAWQGRSADVTGYVSDENRPPVPDTLNP
LIANCIYLLEETIGPLVEEAARAADAQHSAARSILVTEVDLREVERRRRTGIP
ASRAIEAVTQRLKGMPGDPLIHMSWHPLVVOAGMYGRIDLRELEERWSE
CGLOPCWRDAVPVRDPDGTVPWMALPMAHNOIDAWVVYTTSAAFLTSALSGMRS
ELATREFCARAKNSNGRRITYALSWSNDDETGOQLGLELPDEGNFRALRIATAMT
AORPGILAAVLIHLKHLNVATAEGYAARPGGHOQAFAEVAEEAHLRIJYAAVYG
YORGILPISGOKLCILKAGTKPADPELDHEHDAGPTVVIDRVAEVLKAKKKTLLGVC
NYCWSDSKACILCLAGTGPDADELDMCSASCAPDATHSQRQIMADHADTOXAV
FLGNPRLSKPPRARONAFERRATRIVAIEDAAGHPVELHS"
complement(1042..1045)
RBS complement(1042..1045)
/gene="SCD19.03c"
gene complement(3136..4911)
/gene="SCD19.04c"
CDS complement(3136..4911)
/gene="SCD19.04c"
/note="SCD19.04c, hypothetical protein, len: 571 aa; similar to some plasmid derived hypothetical proteins, e.g. TR-085872 (EMBL:AF079317) Splingomonas armaticivorans hypothetical 65.3 kDa protein (plasmid pNL1), 595 aa; fasta scores: opt: 344 z-score: 399.6 E(): 9.4e-15; 26.7% identity in 561 aa overlap. Contains two TTA leucine codons, possible targets for bldA regulation"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

		/protein_id="CAC08305.1"
		/db_xref="GI:10129750"
		/translation="MPAEIQPLAVLFTLPDGVSHGHVLDLPNQQLAVALDLAVGLVAAT
		THHPGIFRHSVARQIYMTTRRMARVDLAAGFTGGADLTLPATLVQYTWLCVDHREER
		IIRVLNFOETRSAGQIDPGICRHLAGRIINVEYSRPQPYSDGEMLRLASCTNOIAA
		AARROALPEAAGADPSSHGVTEDMIAMLIWJACGAERRAERGRGAVGOOG
		AAVAALFPSTEFAYLTFLPMFRTGDIVPGIDVLRIDNTITRSANALLSYRGREG
		NEMILPRDAVRLLDRLELSAOLRDHAAGLABRLMYVCCDIGRCROGTIPDPREGO
		KORAMHESSGVLEDGQPIPVHGGRVRRATYHHRRDTANTGRTIDPNISAREGCH
		ILSSHTEAQILDLEGITIEQAQGDVRRKRAAPVVAYSDEALAFADDFPHVEDGDLDA
		AVSALLSGEODMEVNAACSPYNGAHAPGLCPARPVCLLCPLAAAPRHLPLMLRL
		KEVFSRQAQQMTGQFLRIFGPYPYARLDEVDLPRFGPAALIEATROSATFATPLEPLH
		EEOQP"
RBS		complement(3210..3214)
		/gene="SCD19..04c"
gene		complement(33979..33981)
		/gene="none"
misc_feature		complement(33979..33981)
		/gene="none"
		/note="TfA leucine codon. Possible target for bldA
gene		regulation"
		/label=TfA
		complement(33991..33993)
		/gene="none"
misc_feature		complement(33991..33993)
		/gene="none"
		/note="TfA leucine codon. Possible target for bldA
		regulation"
		/label=TfA
gene		complement(4911..6263)
		/gene="SCD19..05c"
CDS		complement(4911..6263)
		/gene="SCD19..05c"
		/note="SCD19..05c, possible integrase, len: 475 aa; highly
		similar to FR:09RAJ3 (EMBL:AJ250372) Mycobacterium sp. CP1
		putative integrase intM, 451 aa; fasta scores: opt: 2152
		Z-score: 2497.3 E(): 0; 72.8% identity in 441 aa overlap
		/codon_start=1
Query Match	37.4%; Score 374; DB 3; Length 22991;	
Best Local Similarity	63.5%; Pred. No. 3.le-32;	
Matches 610; Conservative %	0; Mismatches 335; Indels 15; Gaps 2;	
OY	42	cggggaagaatctgcgtccacggttcgcagacgcgctcgctcagaggatcatcgagacct 101
Db	7999	CGGGGAAGAATCTCGGCCAGAGACGCACACACTCTCTCCTCACTCAGCTCAAGCGC 7940
OY	102	tcaagcgcgagtgtagcgcgtcgcggagtctcgcgggcgatgaagcgcgacgtgtccgg 161
Db	7939	GCAAGCGCGAGTCACAGTGCTCTCGTCGTCGCGGGGCCCAAMAATGGCGGTCACGCCG 7880
OY	162	gccgggaaagacaaccatgtagaccaccaactcggcgcggttcttcgcagagttcgcgaca 221
Db	7879	GCCGGGTAGAGCGCACGCGCCGCTCGCCGCGGTGACAGCGCGCTTGTCGAGCA 7820
OY	222	acttcctgtagccctcgatccggcgcgacgaggaagcgaacagcacctcgcgacggccct 281
Db	7819	GGTCTCTGATAGCCCTGGATTGCTGCTGCCGCTTGIGTCTCTTAGGCCATCGCAGCGC 7760
OY	282	gcgcgcacttcaacgcggtgcccgcgcgcacattcttcacgcgtctcgcttgagaccgcgc 341
Db	7759	GACCCTCGCCCCCGGCGCCCTGTGTGCGCTCTTCTTTGATACAGCGCCGCGACGACCGC 7700
OY	342	cgtgacgagcgagaccaaagcgaaagaccgtaagcctctgcygcgagcgacacct 401
Db	7699	CGTGACAGCGGCGACACCGGATGACCCCGACCCCGTAATCCCGCGCGCGGAGATCACT 7640
OY	402	ccaagtcgagcttcgcgagaccgcaggttgtatacagcaatgtgtggaagacatatcccaagg 461
Db	7639	CCATCTCGGCGCGCGCTCGGCGCAAGTTGTACAGGCACTGCTCGTACACAGCCGATGG 7580
OY	462	agtg---cgcgcgagcggttctctctctcgcgcgcgatgtgtccagcccgcgaaagtgc 518





gene  
CDS  
/product="hypothetical protein SCD12A.03c"  
/protein\_id="CAB93394.1"  
/db\_xref="GI:8249973"  
/translation="WRTAVRRRAATVLTSTAALLAFAAEAAAGVIGDGGITLL"  
complement(1987..4590)  
/gene="SCD12A.04c"  
complement(1987..4590)  
/note="SCD12A.04c"  
similar to: TR:053874 (EMBL:AL022004) Mycobacterium  
tuberculosis hypothetical 79.7 kd protein (fragment)  
MTV043.55c aa; fasta scores: opt: 651 z-score: 618.2 E():  
5.7e-27; 36.1% identity in 886 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SCD12A.03c"  
/protein\_id="CAB93395.1"  
/db\_xref="GI:8249974"  
/translation="MSTEEKSAAPRSALAEALVRDDVSLAALLRSRDLITPVDTTLT  
LPATAGRSASVVRALREDFRLPOTAEALLAVDDPASYSGLLALMGDDGDDPVA  
LPRAALLREQALVWGADDRLLRLVTRTAELLAPSPQHSPTGLLIVRETRAGSPAA  
IODIIAALGSPSTHDVAVSALSLSLEFDRRRMAALLAEIPWESREVLDTLVWSPPIG  
VHTPPAHRLALDLRGLLPLTPATVTVLPREVALHLNAGANHAPEPVDVPAAT  
HPPQVDPAAGQALAAATATDELKEMDEGGPTVLRAGLSVDLKRITAAIDVPEP  
VAFWELAVAGGLIASGDEAEERYAAVPADEMRLEPAPRRMLAGTWTARTPQG  
VWGDAADRTISALGPNLIDRSAAPEVVRVALLAGLPEASPAVSATVLTARPEEP  
LRGPOQARTGAGARREDLRSRTAKWTLSSBELCTGRCALAAPGALIGAPEAPR  
PATNADTVGGPGDKLPVHHNRTPPVATPTPAERAAATVAVRLALAPPEPLDHV  
LQADVLAVAPGPEGRGLADVLADVSESGATVVRFTGVSRLADMGOSADHLH  
AFILASHTSPVQPLTYLIDVARRHGRLRGASAAVYRCDDDTDEILADKRAAGL  
GLRLTAPVLAAGADPALLDLGLRAGIAPAAESAGVYLTAARDSHRTPAPEPV  
PDGPAPDDTILAAIRAVRAGDLASTPRKRGDGDGSGSGMPCGPIPRCAATET  
LATMQAAVTEGALWIGTVNAGASQNVAPIRVEGFTVADHTADEVTYPLHVR  
TGVAEALADAG"  
4640..5740  
/gene="SCD12A.05"  
4640..5740  
/gene="SCD12A.05"  
/note="SCD12A.05, possible integral membrane protein, len: 366 aa. Contains possible hydrophobic membrane spanning regions at C-terminal domain. High content in alanine, arginine and glycine amino acid residues"  
/codon\_start=1  
/transl\_table=11  
/product="putative integral membrane protein"  
/protein\_id="CAB93396.1"  
/db\_xref="GI:8249975"  
/translation="MRCRRPGGGRGTGVRREGTGHGSPQRTAPEGVNGIESOVVFE  
EYLSVGVGAOQRLPSTATRMILVSELNEIDRRIRAKTTVSPAARIRIDRLDDPAD  
IVTAAGVSGVROQAPSPSTVAPPEQRDGPDPDAARRGRSPAEETPAAPPAASPPHIL  
ASARELGGVGDQDMWRVDDGPFAGGGLGDSVGVGVGVEIPEILKRPQAVDEEER  
TEPAPRPVEEAEPTAGVAKRAARALPLMEGQGNPNLLILAAALLTGVGVASSLS  
LIPILGMILVLSRRLTPAOSKMAVIGLPTGVAAAGVYLMGTEGKWQDPVAEGHIM  
NDATIEHPWVILRTAAVATALYLKMRSRPHG"  
5832..6186  
/note="6x degenerate repeat:  
CCGG(T/A)CCG(C/G)G(A/C)(G/C)"  
6293..6955  
/gene="SCD12A.06"  
6293..6955  
/gene="SCD12A.06"  
/note="SCD12A.06, possible hydrolase, len: 220 aa; similar  
to: TR:086530 (EMBL:AL031124) Streptomyces coelicolor  
putative hydrolase SCIC2.30, 244 aa; fasta scores: opt:  
128 z-score: 155.4 E(): 0.34; 32.9% identity in 167 aa  
overlap. Contains Pfam match to entry Pf00702 Hydrolase,  
haloacid dehalogenase-like hydrolase"  
/codon\_start=1  
/transl\_table=11  
/product="putative hydrolase"  
/protein\_id="CAB93397.1"  
/db\_xref="GI:8249976"  
/translation="MASTTRPAPVGVFDLMTLIDRSFGIRACVVELSARTGYVAD

Query Match	34.7%	Score 346.4	DB: 3	Length 30657
Best Local Similarity	65.5%	Pid. No. 2.1e-25		
Matches 544	Conservative 0	Mismatches 271	Indels 15	Gaps 2
misc_feature	6314	.6820		
gene	/gene="SCD12a.06"	/note="Pfam match to entry PF00702 Hydrolyase, halolacid dehalogenase-like hydrolase, score 28.00, E-value 3.5e-05"		
CDS	/gene="SCD12a.07c"	complement(6904..7182)		
	/gene="SCD12a.07c"	complement(6904..7182)		
	/note="SCD12a.07c, possible integral membrane protein, len: 92 aa. contains possible hydrophobic membrane spanning regions. High content in alanine amino acid residues"			
	/codon_start=1			
	/transl_table=11			
Db	23377	GGGCCCCGGCCCCGGGAAGACTCTTCAGACCCGGCCAGCACTCCCGCAGACTCCAG	23436	
Qy	31	ggaagccacacgaggggaagatctgtccatgctcagagacagcgactcgtccagagatc	90	
Db	23437	CTCCACGGCGCCGACAGGCGGAGCGAGCGTCTGGCGGGGCGGCGCCGACGATGGGGCC	23496	
Qy	91	cgcgagacgcttcagcgcgcgagatctgactctgcggggttcgcgggcagatgaacgcgc	150	
Db	23437	CTCCACGGCGCCGACAGGCGGAGCGAGCGTCTGGCGGGGCGGCGCCGACGATGGGGCC	23496	
Qy	151	ggcgaatgcgcgggcgcggagacgacccatgacagcccaactcgcgcgggtcttcgcgag	210	
Db	23497	GGTCAACGGCGGGCCGTGGTACAGACGACCCGACGCGCGCTTCGCCGCTGCACGCCGTG	23556	
Qy	211	gttcgcgcagaactctctgtatgacctcatgcgcgggcgcgcagaggaacgcgaacagactg	270	
Db	23557	CTTGTTCAGACAGGTCTCTGTACGCTTGATCTGGCCCGCTTCGCCGGGTGGCCAGCGCC	23616	
Qy	271	cgcacgcgcctgcgcgcgacttcaacgcgcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgc	330	
Db	23617	GTCCGCGCGCGCTCCGGAGCGCCGCGCGACCGCGCTTCGTCGTCGTCGTCGTCGTCGTC	23676	
Qy	331	gagcagagccgcgcgtgc	390	
Db	23677	GAGCAGCCCGCGCTTCGACGCGGAGCACGAGGATGACGAGCGCGTACTCTTCGCGCGCC	23736	
Qy	391	gggcgcgcacttcagactcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	450	
Db	23737	CGGGATCACTCTCACTCTGGCGCGCGCTCGCGCCAGGTTCAGACGACCTCTCTCTAC	23796	
Qy	451	catgcccaggg---agtgcgcgcgcgcgcggtcttcctgcgcgcgcgcgcgcgcgcgcgcgcgc	507	
Db	23797	CAGGCGGATGGTGGCGCGCGCGCGCGCGCGCGCGCTCTCTGTGGCTTCGCGGATCTTGTACC	23856	
Qy	508	cgcgaagcttcgc	567	
Db	23857	GGGGAAGTTTCAGACGACCCCGCTACAGCACCCTGGCCGTGGAGACAGAGGTGTGATGCG	23916	
Qy	568	ctgcacacacttcgtccacgc	627	
Db	23917	CTGCGCAGATCTCTGTCGAGAGCGGCTGGCGCGGTGTCATGTGGTGAACCTGTGAGAGTGCAT	23976	
Qy	628	gtgtgcgcagccacgc	687	
Db	23977	GTGTGCGGTGTGAGACCCCTTGAAGCTGTGACCTGACCCGACCGCGAATGTTTCACGCGCGGA	24036	
Qy	688	cagccgcgtctgcct-----gaagcgccgcgtcatcctgcgcgcgcgcgcgcgcgcgcgcgcgc	735	
Db	24037	CAGCTTGTCTGTGTTGGGCGACGCGCTCGCGCGTGGCGCGCCATGTTGGCCGTACACTTGGT	24096	
Qy	736	cgcgcgcgcgtgtctcgc	795	

[illegible]

```

gene
    protein_id="AAE84536.1"
    /translation="MRGVYVLEAKCTAARKVRPMPVLGPNALIVRPRLATPTACTDVMH
    IATPALBRTVKVIGHEAVYVEKYEVENKAFPRGGDRYLPCGSMDRRHPRAQGEANT
    YOTNSPIIDSDPSIGVFSEYVRTLDLADLARIPDKVSVDLVAMLPDMARAGTGTVE
    RLDIQTETELVALMGDLGPVGIMAAAGAALKRAGRNIQDSRPNPLVELRADVIDDFE
    TGGOTTKEKVALLTAGOPDVSVILFASGSTASEFTTLAKAVFGGRACVSGFLFDSEVV
    GTILEWNVNGMEKFLETGVHVODGRDRIERLLAMECGRFDPRLVISHLVHWDSLQO
    GIELRSRDPAVIKRPIVTI"
    complement(4813..6156)
    /gene="XFI728"
    /complement(4813..6156)
    /note="similar to GI|290511 (percent identity: 27 %/query
    alignment coverage: 84.8 %/subject alignment coverage:
    84.0 %) : identified by sequence similarity; putative; ORF
    located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="transport protein"
    /protein_id="AAF84537.1"
    /db_xref="GI:9106796"
    /translation="MNCPARPCCGCCSRBRALFTLPPLSAPRAWICROIARGHQO
    AVSTLTHTPIPASSABADTLPTPALMLMTGFIAVTTETLPGILLPOIADLATIPA
    RAGOMYTVTAAGSLLAIPVLSLTQRKKPYLMIALAEFLFNITATSPWPALTU
    VAFGLVASGSLMGLLGYSARRMYVDIRLRGAALVAMOTPIALSIGVLFETTGTGI
    IGMRGAFVSLVAAILLVSMARMPEVPQEPERRISISVLAGSAGVSIILTALMW
    VTMOYMLYTIYAVEAGTILGRADVDDLILAEGVAAYIGIMVAGNRHRIRLHYIGSL
    VLNAVATVPTSPASALEFAVPMMMLMGASFSGASTSIORTAASDAAGEVDYGAMLT
    TWNAGIAAGCALVTLARGCASVLPASMLPLIVALNALACRRHGFRAGARA"
    complement(6153..7028)
    /gene="XFI729"
    /complement(6153..7028)
    /note="XFI729"
    /note="similar to GI|6459673 (percent identity: 61 %/query
    alignment coverage: 97.9 %/subject alignment coverage:
    87.2 %) : identified by sequence similarity; putative; ORF
    located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="phenylacetaldehyde dehydrogenase"
    /protein_id="AAF84538.1"
    /db_xref="GI:9106797"
    /translation="MKRIDASLSCGFATGDTYNRGCFGMARITGPDWKEPEDHDEA
    IRULKRPEIGVDLIIDTADSYGPSVSOLIDLADLHPGGIKIATKGLVYPKGSTNP
    SMIYVIGDPATLRQCVTMSLKRLKDQLDLMQHRTDKVPRAEDFGAIRFEIDGLIR
    HAWLSQSVAIEEARKEPVAVTYQNUNLNADRADEDVDLYCEANGIGTFIPWEPLAAG
    DLKPAGVADALAKAKCATAGOIALAMLTKRSPVILIPICTSKVAHNENVAAAATLU
    SDEEFALIDAAPRG"
    complement(7165..8055)
    /gene="XFI730"
    /complement(7165..8055)
    /note="XFI730"
    /note="similar to SP|P30864 (percent identity: 35 %/query
    alignment coverage: 96.6 %/subject alignment coverage:
    94.1 %) : identified by sequence similarity; putative; ORF
    located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="transcriptional regulator (LysR family)"
    /protein_id="AAF84539.1"
    /db_xref="GI:9106798"
    /translation="MQVYHLCFSFAAAREQNDPSSVSRAVVALEALTRLFARNTR

Query Match      31.5%; Score 314.8; DB: 1; Length 12335;
Best Local Similarity 58.9%; Pred. No. 8.1e-26;
Matches 594; Conservative 0; Mismatches 387; Indels 27; Gaps 2;

6   acceagcttcgcgcccgcgcggcgaaggcccgagaaatcatcgtccagtccg 65
||||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db  117 ACCAGGCATTAGCCTTGTCGCCCTCCACCACCCGGCGCAAGCCAATCTCTCCACACTTGG 176

-g-66 acagagcctgctcgtgccaggttaccgcggaagccttacgcgagagtgcaagctgctcg 125
```

[illegible]

## SOURCE

regulatory protein; secreted beta-galactosidase; serine/threonine protein kinase; tetr-family transcriptional regulator.

## ORGANISM

Streptomyces coelicolor A3(2)

## REFERENCE

1 (bases 1 to 34074)  
Actinomycetales: Streptomycetaceae; Streptomyces.  
Redenbach, M., Kiser, H.M., Denapate, D., Elchner, A., Cullum, J., Kinash, H. and Hopwood, D.A.

## TITLE

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

## JOURNAL

Mol. Microbiol. 21 (1), 77-96 (1996)

## MEDLINE

97000351

## REFERENCE

2 (bases 1 to 34074)  
Oliver, K. and Harris, D.

## JOURNAL

Unpublished

## REFERENCE

3 (bases 1 to 34074)  
Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

## AUTHORS

Submitted (23-NOV-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge

## TITLE

CB10 15A E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

## JOURNAL

On Nov 26, 1999 this sequence version replaced gi:6468236.

## COMMENT

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>)

CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al.,

Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at

<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream

ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most

upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the imported clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a

small overlap between neighbouring submissions.

Cosmid F81 lies on the Ase-I genomic restriction fragment.

Location/Qualifiers

1..34074

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db\_xref="taxon:100226"

/clone="cosmid F81"

1..118

/gene="SCF81.01"

1..115

/gene="SCF81.01"

/note="nominal overlap with Streptomyces coelicolor ST5B7"

<1..118

/gene="SCF81.01"

/note="SCF81.01, ABC transporter (partial CDS), len: >38 aa: identical to C-terminal region of previously sequenced TR:Q9ZNB0 (EMBL:AB019513) Streptomyces coelicolor ABC

transporter, 584 aa"

/codon\_start=2

/transl\_table=11

/product="ABC transporter"

/protein\_id="CAB61535.2"

/db\_xref="GI:6468413"

/translation="TVVADAGKRVAVGTHRELVTADPLXELATQFLATG"

complement(127..411)

/gene="SCF81.02c"

complement(127..411)

/gene="SCF81.02c"

/note="SCF81.02c, unknown, len: 94 aa"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein SCF81.02c"

/protein\_id="CAB61536.2"

/db\_xref="GI:6468414"

/translation="MISRRIVAVVGLAGVYGLAAPMASAAGGPGTCKTSTNTL

DSLAVGDMRPAEDRAMPLPSQQLNRLLDQLDQVYGLVSPFGVPAIG"

583..587

593..958

/gene="SCF81.03"

593..958

/gene="SCF81.03"

/note="SCF81.03, unknown, len: 121 aa"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein SCF81.03"

/protein\_id="CAB61560.1"

/db\_xref="GI:6468262"

/translation="MGTTRAGPGFDEALRRTGEGTAPPLSLVARDAEVRYDRYD

PSRVLVLRDEIAELMDVYSRDLTHMARCVGGVDAFSECTPDGVRVLSKM

LSLRGELVEQTMQWDE"

complement(986..1645)

/gene="SCF81.04c"

complement(986..1645)

/gene="SCF81.04c"

/note="SCF81.04c, possible tetr-family transcriptional

regulator, len: 219 aa; similar to various Streptomyces

transcriptional regulators, e.g. TR:Q9ZBR0 (EMBL:AL035161)

Streptomyces coelicolor putative transcriptional regulator

SC9C7.10C, 272 aa; fasta scores: opt: 189 z-score: 243.3

E(): 3.5e-06; 29.8% identity in 218 aa overlap. Contains

match to Pfam entry PF00440 tetr, Bacterial regulatory

proteins, tetr family"

/codon\_start=1

/transl\_table=11

/product="putative tetr-family transcriptional regulator"

/protein\_id="CAB61559.1"

/db\_xref="GI:6468261"

/translation="MAGRLRAPTGKRYGKSAQROAERRRRFLDAALQFDGSPGYRA

TTVAALSEAGLSTRQFEERFTEEDVLAALHLQVNAAEVVRPAAGARGPLADR

AAVIRAYANVATDPDRVRITFEIVGVSPRLSEORLARARAVDLCAEDRSAAVR

GEATVRDRRLAATATATIGVNGLHSDYSGWWDATLDEVDELVHLLAVLRAPSPSE

G"

complement(1421..1564)

/gene="SCF81.04c"

/note="Pfam match to entry PF00440 tetr, Bacterial

regulatory proteins, tetr family, score 18.40, E-value

0.0023"

complement(1652..1656)

1832..1836

1845..3098

/gene="SCF81.05"

1845..3098

/gene="SCF81.05"

/note="SCF81.05, possible serine/threonine protein kinase,

len: 462 aa; low similarity to C-terminus SW:PKMA\_THRCU

(EMBL:U23820) Thermomonospora curvata putative

serine/threonine-protein kinase PKWA (EC 2.7.1.1), 742 aa;

fasta scores: opt: 157 z-score: 167.3 E(): 0.06; 23.6%

identity in 309 aa overlap. Contains possible N-terminal

signal peptide sequence"

			/codon_start=1 /transl_table=11 /product="putative serine/threonine protein kinase" /protein_id="CAB6158.1" /db_xref="GI:6468260"
			/translation="MPAFSTRHLGVAALATLTVASPAATAAPADPACILREMEVWG NMMDIADVINSTGDFARIGIDIVDPKAEEMAINADPIKMIFFAALRVGVGHQD LVDMYSSTPDTSVYRSPSPADVSVDLTGDTNMGFRVSGSDMAAYSPGTBRA VSAASTNVHVIDIRTEGEORPSFGTGGPKPHNITFRTGYIMNSICDVTALDAPL DWTGDRRIYVDAKITYROKYAVIDMRRLAIGIGDSADVARPAVEFPEDSKLTFVFS FNFEFEEDVAGDKITRTTKLPKPATSEDTTHWNSRHGLMSPDGARICVAGTM DDVATVNVRDLIEEGPLVPAPKPYAWIASGDGHCYVESGGADVTAIDLETGERSVS VPVDHPQRVIRIGVVADWTGPSAG"
gene			complement(3339) .3839) /gene="SCR81.06c" /note="SCR81.06c, hypothetical protein, len: 166 aa; similar to various hypothetical proteins, e.g. TR_AA036153 (EMBL:AE001768) Thermotoga maritima conserved hypothetical protein TM1076, 157 aa; fasta scores: opt: 293 z-score: 385.9 E(): 4e-14; 39.1% identity in 133 aa overlap"
CDS			/codon_start=1 /transl_table=11 /product="hypothetical protein SCR81.06c" /protein_id="CAB6157.1" /db_xref="GI:6468259"
			/translation="MRLITSDTLPRRAELPAPLAETIPRADVLHAGDWDATLL DLESRLRRVGVGNNDGPALRARLEPVAVADLGFRGVHETGAOGRANCAAR EPGLDILVFGRSHIPWTTATATGLRLINPSPTRRRQPPRTYTATVAAGRILDVEL HRLEPR"
gene			3962.4543 /gene="SCR81.07" 3962.4543
CDS			/gene="SCR81.07" 3962.4543 /note="SCR81.07, possible hydrolase, len: 193 aa; similar to TR:P94573 (EMBL:D82987) Bacillus subtilis hypothetical 21.1KD protein, 189 aa; fasta scores: opt: 504 z-score: 582.8 E(): 4.4e-25; 48.4% identity in 190 aa overlap and to SW:EMTB_ECOLI (EMBL:M24148) Escherichia coli isochorismatase (EC 3.3.2.1) (2,3 dilydro-2,3 dihydroxypentazate synthase) EntB, 285 aa; fasta scores: opt: 205 z-score: 242.1 E(): 4.1e-06; 26.8% identity in 183 aa overlap. Contains match to Pfam entry PF00857 Isochorismatase, Isochorismatase family"
			/codon_start=1 /transl_table=11
Query Match	21.3%; Score 213.2; DB 3; Length 34074;		
Best Local Similarity	55.3%; Pred. No. 2.3e-15;		
Matches	488; Conservative 0; Mismatches 373; Indels 21; Gaps 3;		
OY	26 ccgcccgaagccaccgcggcgaaagatcttcgttcacgttcgacaagcgcctctgcgtcacg	85	
D3	33157 CCGGCGATCGGGGTCGGCGCGCGAGGCCGCCTCCGATCTCCGTCCTCGCTCACG	33216	
OY	86 gtcatcgcggagcccttaagcgcggagcgcggagtctctcgggggttcgcggccaataacg	145	
D3	33217 GTGAGTTCGACCGCCGCCACGCGCTCTCTCGAGAGTGGCCCGCCGGAGAGCGCGACGATC	33276	
OY	146 gcyccgcgatgcgcggcgccggagcacacccatgcgagcccccaactcgcgcgggtcttcg	205	
D3	33277 GCGCGCGTCCACGATGCTCTCGTGGCAGCAGCCAGGCCAGGCCGCGGTCTCCCTCTCGATAC	33336	
OY	206 ccgaagttgcgcgaacttctcgttagccttcgattcgcgcggcgcgaggaagaagcaaacgc	265	
D3	33337 CCGGCTCTCCCCCGCATCCGGGTGACGCGCTCTCCACACGCGCGCGGTCTCCCTCTCGATAC	33396	
OY	266 acctgcacagcgcctcgcgcgcgaacttaacgcggcgcgcgcgagcttctcacgcgt	325	
D3	33397 AGGGTGTGTCGCCAAGGTGTGCGT-----GGCTGTGCGCCCGGTGGCGGTGTGCTCCAGTGC	33450	
OY	326 ccgctgaagcaaggcgcgltgcagcggcgaccaagcgcaagaagcgcgtagcgttcgc	385	

	Db	33451	CGGGTGAGCGGCCGCCCGGCGACGCAGCTCCAGGGCACCAGCCGACTGGTCCGG	33510
OY	386	gcgcgagcgaacacactccagctcggtcgctgcggacgcgcaggatgtgtacagcaattgtg	445	
Db	33511	CAGAGGGGCGAGCATTCGCCGCTCTCTCCGGGGTGAGACCAAGATTGTATGGTTTCGCAATG	33570	
OY	446	gagaccatgcccagagagtgtgagggcgagccggttctcttcctcgcgcgagcatgtgcacg	505	
Db	33571	GACACGAACCTTGTCACAGCCGTCGCCCTCCGCGGTGTACTGCAATCTTGAGAACTGCCAG	33630	
OY	506	cccgagaagttcgaacgacgacgtaggagacacttgcgcgttgcgacgagagctgtcacatg	565	
Db	33631	CGGTACATCGAACACTGGCCCCGATGTAGCGGACCTTGCCGCCCTTAGCCAGGTCGACG	33690	
OY	566	gacctcaaacacttgttccccagggcgcgacacggttcgatgtgtgtacatltgtatacgttcg	625	
Db	33691	GCCCTCATCTGTCCTCTCAACCGGGATGTGCGGGTGAAGAAGCGATGTGTGAAGGTGG	33750	
OY	626	atgtgttcgaacgcgccagcctgtcgacgacatccctctcgacgagagcgatatgtgcgcgcc	685	
Db	33751	ACGTAGTGTGGTGTCCCAGGCGCGCGACGCGTGTGATGATCTGGTCATATCACCGCGCTTGGG	33810	
OY	686	gaacagcccgctgtgtgttaacgagctcgtcatctcgtccgcgcgcagactgtgtgcgaacag	745	
Db	33811	GAGAACTCCGGCGCCCTTTGGGCCGGGGGCGATCCGGCGCTTACCTGTGTGTCGACGACG	33870	
OY	746	gtgtctctcgagcgcgcgcgcggcgacgacacacgttgcacacagctctctcgtgtg	805	
Db	33871	ATGTGCTGCGGGCG---GGCGAAATCGGCGACATGCTTTCGCGAGCATCTCTCGTGGTGG	33927	
OY	806	cccttgttgaagccgcgcgcggtacatgttcgcgcgtgtgtgagagcagtttatgtccgcggtcc	865	
Db	33928	CCGTCGGGA-----GTACACGCTTTGGCGGTGTGAAGAAAGTGAACCGCCGCCCTTC	33975	
OY	866	cgggcgtgttcacatcagcgacgacgctgtctgtctcgtcgaag	907	
Db	33976	AGCGCTCGCGGATACGCGGACGACGACGCGCCGCTCTGTGTGAGG	34017	
RESULT	13			
SCK13				
LOCUS	SCK13	39896 bp	DNA	BCT
DEFINITION	Streptomyces coelicolor cosmid K13.			26-FEB-2001
ACCESSION	AL451182			
VERSION	AL451182.2	GI:33162229		
KEYWORDS	ABC transporter ATP-binding protein; aldoketoreductase; amino acid deaminase; deaminase; dehydrogenase; DNA-binding protein; integral membrane transport protein; marR-family transcriptional regulator; membrane protein; meth. cystathionine gamma-synthase; narg3; nitrate reductase alpha chain; narHJ; nitrate reductase beta chain; nitr; nitrate reductase gamma chain; narJ; nitrate reductase delta chain; oxidoreductase; peptide methionine sulfoxide reductase; regulatory protein; tetR-family transcriptional regulator; transcription elongation factor.			
SOURCE	Streptomyces coelicolor.			
ORGANISM	Streptomyces coelicolor			
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
REFERENCE	1 (bases 1 to 39896)			
AUTHORS	Redenbach,M., Kiese,H.M., Denapalte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.			
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome			
JOURNAL	Mol. Microbiol.	21 (1),	77-96	(1996)
MEDLINE	97000351			
REFERENCE	2 (bases 1 to 39896)			
AUTHORS	Oliver,K. and Harris,D.			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 39896)			
AUTHORS	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-Dec-2000) Streptomyces coelicolor sequencing project,			

## COMMENT

Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
On Feb 28, 2001 this sequence version replaced gi:11691812.

## Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC787.01c. SC (S. coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid K13.

## FEATURES

```

source
    1..39896
        /organism="Streptomyces coelicolor"
        /db_xref="taxon:1902"
        /note="nominal overlap with Streptomyces coelicolor cosmid SC2K31"

misc-feature
    complement(1..109)

source
    1..39896
        /organism="Streptomyces coelicolor A3(2)"
        /strain="A3(2)"
        /db_xref="taxon:100226"
        /clone="cosmid K13"
        /complement(113..430)
        /gene="SKK13.01c"
        /complement(113..430)
        /note="SKK13.01c"
        /gene="SKK13.01c"
        /note="SKK13.01c, hypothetical protein, len: 105 aa; similar to TR:O9K2T7 (EMBL:AL353833) Streptomyces coelicolor hypothetical 20.2 kDa protein (fragment) SC65.01c, 183 aa; fasta scores: opt: 191 z-score: 223.2 E(): 6.8e-05; 41.5% identity in 82 aa overlap"
        /codon_start=1
        /transl_table=1
        /product="hypothetical protein"
        /protein_id="CAC18718.1"
        /db_xref="GI:11691849"

CDS
    /gene="SKK13.01c"
    complement(113..430)

gene
    /gene="SKK13.01c"
    complement(611..1291)

CDS
    /gene="SKK13.02c"
    complement(611..1291)
    /note="SKK13.02c"
    /note="SKK13.02c, possible regulatory protein, len: 226 aa; similar to TR:O9KXX1 (EMBL:AL357432) Streptomyces coelicolor putative transcriptional regulator SCD95A.03c, 199 aa; fasta scores: opt: 414 z-score: 464.7 E():
```

```

misc-feature
    complement(1049..1169)
    /gene="SKK13.02c"
    /note="Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family, score 49.20, E-value 4.1e-11"

RBS
    1400..1404
    1412..2404
    /gene="SKK13.03"

CDS
    1412..2404
    /gene="SKK13.03"
    /note="SKK13.03, probable oxidoreductase, len: 330 aa; similar to TR:AG05923 (EMBL:AE004681) Pseudomonas aeruginosa probable oxidoreductase PA2535, 331 aa; fasta scores: opt: 1001 z-score: 1136.6 E(): 0; 48.2% identity in 328 aa overlap. Contains Pfam match to entry PF00248 Aldo_Ket_red, Aldo/keto reductase family"
    /codon_start=1
    /transl_table=1
    /product="putative oxidoreductase"
    /protein_id="CAC18716.1"
    /db_xref="GI:11691847"
    /translation="MQHRTIGQGLRVASGLGVGMSLAYGPNDEEDGISTIRHAHE LGVDFPDAELLYGGCTGNETLGNVAVNDPREVYLAIRFGEDMTSKLGGVNSRPE NIREVENSRLYRTDRDLPFOHSDPDVAGVAGVGLITPGKYRGLISVGP RYIRRAHAHVTPTALQIERYSTFERVEDEILPVVELIGLVPSPLRGLSGVVR AGEYPPEDMRMRDENGENAYNNAAEQLRKLAIGTTPAQLALAMLAQGGDVA PIPGTYSARLARLEENGAIDAELTEADLARIRREVIPOGSAGSRYPASMAAGPRTD 2021..2329
    /gene="SKK13.03"
    /note="Pfam match to entry PF00248 Aldo_Ket_red, Aldo/keto reductase family, score 21.20, E-value 4.9e-05"
    complement(2502..3377)
    /gene="SKK13.04c"
    complement(2502..3377)
    /note="SKK13.04c"
    /note="SKK13.04c, possible DNA-binding protein, len: 314 aa; similar to TR:O9R167 (EMBL:AL109949) Streptomyces coelicolor putative DNA-binding protein SCU11.39c, 279 aa; fasta scores: opt: 544 z-score: 640.7 E(): 3.8e-28; 40.2% identity in 249 aa overlap. Contains Pfam match to entry PF01381 HTH_3, Helix-turn-helix and possible helix-turn-helix motif at residues 64..85 ("4.67 SD")
    /codon_start=1
    /transl_table=1
    /product="putative DNA-binding protein"
    /protein_id="CAC18685.2"
    /db_xref="GI:13162230"

CDS
    /gene="SKK13.04c"
    complement(3117..3290)
    /note="Pfam match to entry PF01381 HTH_3, Helix-turn-helix, score 25.50, E-value 0.0012"
    complement(3380..3384)
    3534..3537
    3546..4586
```







1D7A0VDDIILVCYTPVPGOGADIAKTAALVADMPDEOVAVQINRECSAGLEAVNIGA  
 MKVRSGFEDVLIVGVGESMRVPMGSDGGAALJDEPTNHTSFVPGIGADILATLEG  
 PSRAVDPAFALBSQOKAKARABEGLEFGKSIVYVAQONGVLLDHEPFIADSTLEGLA  
 ALKPSFEMGQMGFDATALRKYSOVERIEHVTPTNNSGIVDASAMFLGASREGL  
 GKARARIVATVSTDPITMLTGPAPATRKALAGLKVEDIIDLEVEAFASVVMK  
 FMKDMGVAESKVNNGAIAIMGHPLGATGCITLGLDELERRNLRYGATLCVGGGM  
 GIATTIERY"  
 1535..3679  
 /gene="PA1737"  
 1535..3679  
 /gene="PA1737"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable 3-hydroxyacyl-CoA dehydrogenase"  
 /protein\_id="AAG05126.1"  
 /db\_xref="GI:9947714"  
 /translation="MTDAIRYEKGODNVVVLTMDPGOSANTMNGVREAMAKTIARL  
 EAEKEIGAVLTSAKKTPEFAGDINELIKYTKADAPAFYOGILELKGQLRRLLETIGK  
 PVVAINGAALGGWEICLACHHRIALDNPQVLPEVTGLGLPGGGIYVMVRLIG  
 LEKALPYLAEGKKVPAELKAGLITHELDSPHEMLEKAWILANPAKOPWDSAGY  
 KIPGTPASPNVQMLAIPVYLNDKTKGCEPAPERIMCAAVEGAQVDFDQILFAR  
 YFTELTTQVAKNMIGTFMFQLINEINACKSRPOGYPVRAKKGVYLGAAGMGAGIAY  
 SAAAGIEVVLKDVSLAEAKGKAYSARLIDKKVARGHLSAEKRDALFARIYVSEAD  
 FEGCDLIEAVEDEBRLKGYTAEEAKHALDAVVAASNTSTLPIINGLAQAVARPEKFI  
 GHFSPYDKMPLVEITIRGEKTSDETLFARGPYVQIKTPIYVNDSRGFTSRVETG  
 FTNEGIALGEGVSAMIDNCAKQAGMVGPLAISDEVSLMTHTIRNOTADLBAEG  
 KALPHPAFAVIDLVNVEYKRPGKAAGAFDYDPANGRKKHMAELKSREREKDAQISA  
 EDVDRLEFIQAIETVRCVEEGVHSTADANGSFEGIGFAMSGALQFIYQYGLKD  
 FIAQYVLEQGERPEPPALILEKAKGELF"  
 complement(3931..4845)  
 /gene="PA1738"  
 complement(3931..4845)  
 /gene="PA1738"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transcriptional regulator"  
 /protein\_id="AAG05127.1"  
 /db\_xref="GI:9947715"  
 /translation="MKKPDLAGLAFAVAIARERSFRAATLGVPPTLSHTLRELET  
 QUGIRLKRTRNVSPTAGEHLALHPAFEDIAALDSLRNRPETPHGLVIRINAPR  
 NAIGVLAPREFGLADYPGITLEVADGEPANIEYEADEGDAIRLGEDIOHMRANVR  
 GPDRLAIVAPFAPYRRGROGSPGSELPGLHRCIGRKVSSGELYKQWFSKKAISVA  
 MSGPLVDDPRLMLOALADYGIARIEEVAEHLAAGRLERVLADKCAFPFGPHLYI  
 PRRNRHPALSAVINLKRPGEEHPGPA"  
 4997..5968  
 /gene="PA1739"  
 4997..5968  
 /gene="PA1739"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable oxidoreductase"  
 /protein\_id="AAG05128.1"  
 /db\_xref="GI:9947716"  
 /translation="MOHYRLGNSGLKVSRLCGMTYDPPAMPVWLDEBRAPFIRE  
 ALEAGIDFFSADADYTSGESERIIGRAIRDFAORREDIATIKAFPPMSDRPACLSR  
 KHLASVDSARLCTDVLDFVYHRRPDPPIETGCTLDSIYAVGVRYIGASMP  
 AMREKMLAFORHGHGLOFISQSOYNLIVRDEEDVLPCREBGIALTTPKSPILARGL  
 LAGARSAGTLRTIDEQAPRMVGGREVESTLGALEKLAARGLPPAQOLAIAWLIGRN  
 GVAAPYVGLSRPHLEDAIAALTDLAEECATLEAPLTARARTPRW"  
 5991..6989  
 /gene="PA1740"  
 5991..6989  
 /gene="PA1740"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG05129.1"  
 /db\_xref="GI:9947717"  
 /translation="MMRQASVLLIVSVSATVLFVARADGEREPCSAANSIYLPAG  
 ARYPNGIAHADGTLVGLVTSGLRLKRPGGEWETPFAGSPALFAAALRLDEREGL  
 LMGNSPDLFPAGRRRPHGVFALDASGAVRRLTLPJGDMGMDLAVADPGLVYISETR  
 DGLMLRLRGPAPFOVLLRDSRLAGPGIGAAAGIYRVDDDTLLVNGPSGAIHAVEGV  
 ATRRLREMLPRLTENPDGMAIAPDGLIYLENAIRSGACKVLRIAEPLAPGRGIE

VIREGLESPPVNLTVTGQCAVYSESRIRHRLLPGRGEVPAERFLIHRPLDAAGR"  
 complement(6998..7369)  
 /gene="PA1741"  
 complement(6998..7369)  
 /gene="PA1741"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG05130.1"  
 /db\_xref="GI:9947718"  
 /translation="MLKAVRNGLCOAIIADPVSAPRKIORSAAEOAVERALNLS  
 LYQFACPCVKTERRAMRLNLPMLQKMDNPPQARQALLBEGGKVKVCLNIEENG  
 VVMWYESSRITVILEGRANA"  
 7614..8336  
 /gene="PA1742"  
 7614..8336  
 /gene="PA1742"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable amidotransferase"  
 /protein\_id="AAG05131.1"  
 /db\_xref="GI:9947719"  
 /translation="MSLRICILETPIILPELIEREYEGYGRMFQOLFAKOPIAEFVYI  
 NYVEGRPADDCSPAYLVYTSKADSPGDPPIOTLKTFLDRYERGXLLGVCEHQ  
 ILALLGKAERASOGWGIHQYQUNERADMSALPDLTLLISHQDQVTLPENAR  
 VIASSDFCPYAAVAGDVLCPGHPHEVHDYSRLLELRQOHLGEELYRKGVDSLAR  
 HQGQIVAEWMMRPFVAPQPTGA"  
 complement(8394..8633)  
 /gene="PA1743"  
 complement(8394..8633)  
 /gene="PA1743"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG05132.1"  
 /db\_xref="GI:9947720"  
 /translation="MRQTRKRWIYALAIILNGVSLGLAGSAGALARTIECQAM  
 VVALANLALKQSGSTPPAPKQRMWIPERFILDCPA  
 complement(8661..8879)  
 /gene="PA1744"  
 complement(8661..8879)  
 /gene="PA1744"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG05133.1"  
 /db\_xref="GI:9947721"  
 /translation="MKRSPRHVARNTALATGLVAILMLVVLIVAGASARNATPD  
 NDTATFSSNOIOALPGLISGINSRRAG"  
 complement(9034..9522)  
 /gene="PA1745"  
 complement(9034..9522)  
 /gene="PA1745"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG05134.1"  
 /db\_xref="GI:9947722"  
 /translation="MPKWTALCLIPILASCATCRDSDREKLLYENDMLQAKLIIVAN  
 DDRRLPLAQLLEBARADRGSEADPFQAVILARRDGOASDCERTTILIERSEARRYP  
 LANMLIKTIVPEFLPPRADIPKATYEQAYAEIVANSGYTSFERAQVYDRLAPT  
 SS"  
 9743..10249  
 /gene="PA1746"  
 9743..10249  
 /gene="PA1746"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG05135.1"  
 /db\_xref="GI:9947723"  
 /translation="MIREVEGGIILISGAOVIAGHTAPQDHFDSGLALALREBWPBMVR  
 DYRAHAHARAPFGSIWAGVDEGKTCQIYNLITQGLHSGRSKAKFGKASLEDVGH

ALRELARYRSEGVSLALPCVATGVGLDMSEKPLVVRHGLDEIPVILYEVYRK  
 VAAEKLIA\*  
 BASE COUNT 1741 a 3254 c 3569 g 1750 t  
 ORIGIN

Query Match 18.5%; Score 185.2; DB 1; Length 10314;  
 Best Local Similarity 52.6%; Pred. No. 4,4e-12;  
 Matches 459; Conservative 0; Mismatches 398; Indels 15; Gaps 2;

```

  34  ggcacccgggggaagatctctcagcttcgagacagcgccgctccgagcgtcagcgc 93
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  550  ggcggcgccgctgacgacgcccctccacagctgacccactcccttcggcgagctcagcgt 581
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  94  ggcacccctcagcgagagtcagctcagcggttcgagcgagcgaatgaacgagcgagc 153
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5890  cagcgccgagcgagcgagctcagcagctgagcagcagcagcagcagcagcagcagc 5831
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  154  gatcgagcgagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5830  gacacccctccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 5771
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  214  ggcgacgaactctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5770  agcgcgagcagcttcacacgacgacgacgacgacgacgacgacgacgacgacgac 5711
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  274  acgagcctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5710  ggcggggggcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 5651
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  334  cagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 393
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5650  cagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5591
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  394  cagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 453
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5590  cagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5531
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  454  gccacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 513
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5530  ttggcgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5471
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  514  gttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 573
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5470  cgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5411
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  574  caactctccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 633
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5410  gcttctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5351
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  634  gacgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 693
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5350  cgtacgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5291
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  694  gctcgtctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 753
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5290  gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  754  ggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 813
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5230  ggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  814  gacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 873
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5173  -----ATAGATATCGGACATCGAAGATCGATCCCGGCTTCCACGCGCTC 5126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  874  gttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 905
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  5125  GCGAATGAGAGGAGCGCGCGCTTCCCTCTCCCA 5094
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15  
 AP03003/c

LOCUS AP03003 349116 bp DNA BCT 03-FEB-2001  
 DEFINITION Mesorhizobium loti DNA, complete genome, section 10/21, complete  
 sequence.  
 ACCESSION AP03003 BA000012  
 VERSION AP03003.1 GI:11994978  
 KEYWORDS HTG.  
 SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.  
 ORGANISM Mesorhizobium loti  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Phyllobacteriaceae; Mesorhizobium.  
 1 (sites)  
 REFERENCES  
 AUTHORS Kaneo,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,  
 Watanabe,A., Ideesawa,K., Ishikawa,A., Kawashima,K., Kinura,T.,  
 Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,  
 Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,  
 Takeuchi,C., Yamada,M. and Tabata,S.  
 Complete genome structure of the nitrogen-fixing symbiotic  
 bacterium Mesorhizobium loti  
 JOURNAL DNA Res. 7 (6), 331-338 (2000)  
 MEDLINE 21082930  
 REFERENCE 2 (bases 1 to 349116)  
 AUTHORS Kaneo,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2000) Takakazu Kaneo, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research, Yana  
 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:kaneok@kazusa.or.jp,  
 URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935,  
 Fax:81-438-52-3934)  
 FEATURES  
 Location/Qualifiers  
 source 1..349116  
 /organism="Mesorhizobium loti"  
 /strain="MAFF303099"  
 /db\_xref="taxon:381"  
 BASE COUNT 63710 a 112865 c 109235 g 63306 t  
 ORIGIN

Query Match 18.1%; Score 180.8; DB 2; Length 349116;  
 Best Local Similarity 52.8%; Pred. No. 1.4e-12;  
 Matches 471; Conservative 0; Mismatches 402; Indels 19; Gaps 3;

```

  34  ggcacccgggggaagatctctcagcttcgagacagcgccgctccgagcgtcagcgc 93
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151381  gacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 151322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  94  ggaagccttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 153
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151321  cgtcggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 151262
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  154  gatcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151261  cagc-aagcgttgacgagcagcagcagcagcagcagcagcagcagcagcagcagc 151203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  214  ggcgacgaactctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 273
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151202  cttagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 151143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  274  acgagcctcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151142  gaaattgaagacgacgacgagcagcagcagcagcagcagcagcagcagcagcagc 151083
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  328  gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151082  gaaagacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 151023
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  388  ggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  151022  gacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 150963
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  448  gacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  150962  ctgcgaagcggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 150903
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:23:55 ; Search time 28.5 Seconds

(Without alignments)  
1545.880 Million cell updates/sec

Title: US-09-463-705A-2

Perfect score: 1728

Sequence: 1 MTTDATATVRLGRSALLTSR.....LDEIFPAVSGAAPPAAWLO 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_16:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.podent:\*  
13: sp.unclassified:\*  
14: sp.vertebrate:\*  
15: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	333	2	033936
2	1415.5	81.9	330	2	09F833
3	1286.5	74.5	329	2	09XC70
4	1056	61.1	347	2	09S0M9
5	1025	59.3	323	2	09F825
6	841	48.7	330	2	09F825
7	804.5	46.6	329	2	09PCO1
8	791.5	45.8	362	2	09PCO5
9	763	44.2	316	2	09PCO5
10	745	43.1	354	2	09KX18
11	495	28.6	323	2	044328
12	472.5	27.3	323	2	09I228
13	428.5	24.8	336	2	09I228
14	421	24.4	307	2	09RUC4
15	414.5	24.0	352	2	09K057
16	408	23.6	316	2	09HXR2
17	399	23.1	324	1	09HPH9
18	393	22.7	344	3	059826
19	387	22.4	343	1	09HI90

20	384	22.2	315	2	09RS10	09rs10 delinococcus
21	380.5	22.0	328	10	023016	023016 arabidopsis
22	379.5	22.0	328	10	049847	049847 egeria dens
23	376.5	21.8	328	10	039151	039151 arabidopsis
24	373	21.6	345	2	09HXY1	09hxy1 pseudomonas
25	371.5	21.5	342	3	002895	002895 saccharomyces
26	363	21.0	419	4	014722	014722 homo sapien
27	363	21.0	419	6	09XTJ31	09xtj31 oryctolagus
28	360.5	20.9	330	10	082064	082064 solanum tub
29	359.5	20.8	408	6	028528	028528 mustela put
30	356	20.6	401	11	063277	063277 mus musculus
31	355	20.5	327	2	09RS66	09rs66 delinococcus
32	344	19.9	401	13	09PWR1	09pwr1 gallus gall
33	340	19.7	319	2	09WYR9	09wy9 thermotoga
34	339.5	19.6	336	1	034185	034185 halobacteri
35	338.5	19.6	351	3	09P702	09p702 schizosacch
36	338	19.6	336	1	09HOU7	09hou7 halobacteri
37	337.5	19.5	387	2	09PGD5	09pgd5 xyella fas
38	333	19.3	453	10	P93818	P93818 arabidopsis
39	331	19.2	367	13	09PTM5	09ptm5 xenopus lae
40	330	19.1	367	6	09NOV9	09nov9 oryctolagus
41	329.5	19.1	330	2	09EMF0	09emf0 streptomyces
42	329	19.0	344	2	09XSG7	09xsg7 streptomyces
43	328	19.0	367	6	027955	027955 bos taurus
44	328	19.0	367	11	064284	064284 mus musculus
45	325.5	18.8	348	2	09EMJ0	09emj0 streptomyces

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT;	333 AA.
033936				
ID	033936:			
AC	033936:			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DF	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)			
DE	ERYBII.			
GN	ERYBII.			
OS	Saccharopolyspora erythraea (Streptomyces erythraeus).			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Pseudonocardiales; Pseudonocardaceae;			
OC	Saccharopolyspora.			
OX	NCBI_TaxID=1836;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL 2338;			
RX	MEDLINE=98015410; PubMed=9353926;			
RA	Summers R.G., Donadio S., Staver M.J., Wendt-Pienkowski E.,			
RA	Hutchinson C.R., Katz L.;			
RT	*Sequencing and mutagenesis of genes from the erythromycin			
RT	biosynthetic gene cluster of Saccharopolyspora erythraea that are			
RT	involved in L-mycarose and D-desosamine production.";			
RL	Microbiology 143:0-0(0).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NRRL2338;			
RA	Salah-Bey K., Domuth M., Michel J.M., Haydock S., Cortes J.,			
RA	Leadlay P.F., Raynal M.C.;			
RL	Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; U77454; AAB84068.1; -;			
DR	EMBL; Y14332; CAAT4709.1; -;			
SQ	SEQUENCE 333 AA; 35867 MW; 6F912355F21B6F8F CRC64;			

Query Match 100.0%; Score 1728; DB 2; Length 333;

Best Local Similarity 100.0%; Pred. No. 3.1e-14;

Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	MTTDAATVRLGRSALLTSRLMTGVNFSGRVEDDARLMDHARDGNCIDTADMGW	60
DB	1	MTTDAATVRLGRSALLTSRLMTGVNFSGRVEDDARLMDHARDGNCIDTADMGW	60

QY 61 RLYKHTTEELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRRL 120  
 DB 61 RLYKHTTEELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRRL 120  
 QY 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIIAAOENARRHS 180  
 DB 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIIAAOENARRHS 180  
 QY 181 LGWVSHOCLNLAVRHAELVLPAAQAYGGLVFAWSPPLHGGLLSGALEKLAAGTAVKSAQ 240  
 DB 181 LGWVSHOCLNLAVRHAELVLPAAQAYGGLVFAWSPPLHGGLLSGALEKLAAGTAVKSAQ 240  
 QY 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLSRPGIAGAVIGPRTPEQDLSALK 300  
 DB 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLSRPGIAGAVIGPRTPEQDLSALK 300  
 QY 301 ASAMTDEQALSELDEIFPAVASGGAPEAWLQ 333  
 DB 301 ASAMTDEQALSELDEIFPAVASGGAPEAWLQ 333

## RESULT 2

QY 09F833 PRELIMINARY: PRT: 330 AA.  
 ID 09F833;  
 AC 09F833;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)  
 DE TDP-4-KETO-6-DEOXYHEXOSE 2,3-REDUCTASE.  
 GN MEGDVII.  
 OS Micromonospora megalomicea subsp. nigra.  
 OC Bacteria; Firmicutes; Actinobacteridae;  
 OC Actinomycetales; Micromonosporinae; Micromonosporaceae;  
 OC Micromonospora.  
 NC NCBL\_TaxID=136926;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-NRRL3275;  
 RX MEDLINE=20430101; PubMed=10972798;  
 RA Volchegursky Y., Hu Z., Katz L., McDaniel R.;  
 RT "Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in *Saccharopolyspora erythraea*.";  
 RL Mol. Microbiol. 37:752-762(2000).  
 DR EMBL: AF263245; AAG13914.1; -  
 SQ SEQUENCE 330 AA; 35434 MW; 32DA7C2DDEF60997 CRC64;

Query Match 81.9%; Score 1415.5; DB 2; Length 330;  
 Best Local Similarity 81.4%; Pred. No. 2.2e-100;

Matches 267; Conservative 26; Mismatches 34; Indels 1; Gaps 1;

QY 2 TTPA-ATHVLRGSAITLSRLMTGTVNFGSRVEDDQALRLMDHARDGICLDTADMYGW 60  
 DB 3 TTGAGSRVAVGSAITLSRLMTGTVNFGSRVEDDQALRLMDHARDGICLDTADMYGW 62  
 QY 61 RLYKHTTEELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRRL 120  
 DB 63 RLYKHTTEELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRRL 122  
 QY 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIIAAOENARRHS 180  
 DB 123 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIIAAOENARRHS 182  
 QY 181 LGWVSHOCLNLAVRHAELVLPAAQAYGGLVFAWSPPLHGGLLSGALEKLAAGTAVKSAQ 240  
 DB 183 LGWVSHOCLNLAVRHAELVLPAAQAYGGLVFAWSPPLHGGLLSGALEKLAAGTAVKSAQ 242  
 QY 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLSRPGIAGAVIGPRTPEQDLSALK 300  
 DB 243 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLSRPGIAGAVIGPRTPEQDLSALK 302  
 QY 301 ASAMTDEQALSELDEIFPAVASGGAPEAWLQ 328

DB 303 AAEITLGEELRELEALFPAPAVDGPVP 330

## RESULT 3

QY 09XC70 PRELIMINARY: PRT: 329 AA.  
 ID 09XC70;  
 AC 09XC70;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)  
 DE NDP-HEXOSE 2,3-ENOYL REDUCTASE TYLCII.  
 GN TYLCII.  
 OS Streptomyces fradiae.  
 OC Bacteria; Firmicutes; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 NC NCBL\_TaxID=1906;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-T59235;  
 RA Bate N., Butler A.R., Smith I.P., Cundliffe E.;  
 RT "The mycarose-biosynthetic genes of *Streptomyces fradiae*, producer of tylosin.";  
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF147704; AAD41821.1; -  
 DR HSSP: P14550; 2ALR.  
 DR InterPro: IPR001395; -  
 DR Pfam: PF00248; also ket. red. 1.  
 DR PRINTS: PR00069; ALDKETREDTASE.  
 SQ SEQUENCE 329 AA; 36144 MW; 14BIE58076AEC09B CRC64;

Query Match 74.5%; Score 1286.5; DB 2; Length 329;  
 Best Local Similarity 74.7%; Pred. No. 1.6e-90;

Matches 242; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

QY 8 HVRLRGSALITSLRLMTGTVNFGSRVEDDQALRLMDHARDGICLDTADMYGW 67  
 DB 5 YVGLGATLVSRMLTGTGTVNFGSRVEDDQALRLMDHARDGICLDTADMYGW 64  
 QY 68 EELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRRLGVPHIV 127  
 DB 65 EELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRRLGVPHIV 123  
 QY 128 YQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIIAAOENARRHS 187  
 DB 124 YQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIIAAOENARRHS 183  
 QY 188 CLYNLAVRHAELVLPAAQAYGGLVFAWSPPLHGGLLSGALEKLAAGTAVKSAQ 247  
 DB 184 CLYNLAVRHAELVLPAAQAYGGLVFAWSPPLHGGLLSGALEKLAAGTAVKSAQ 243  
 QY 248 PSIRPAIEAVEKFCRNLDGEDPAEVGLAWLSRPGIAGAVIGPRTPEQDLSALK 307  
 DB 244 PSIRPAIEAVEKFCRNLDGEDPAEVGLAWLSRPGIAGAVIGPRTPEQDLSALK 303  
 QY 308 EQALSELDEIFPAVASGGAPEAWLQ 331  
 DB 304 EEBHRELEALFPVVGSGGEVPEAWLQ 327

## RESULT 4

QY 09S0N9 PRELIMINARY: PRT: 347 AA.  
 ID 09S0N9;  
 AC 09S0N9;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, last annotation update)  
 DE TDP-4-KETO-6-DEOXY-L-HEXOSE 2,3-REDUCTASE.  
 GN AVEBVIII.  
 OS Streptomyces avermitilis.  
 OC Bacteria; Firmicutes; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.



```
Db 123 YIDVGFHHIDRSTPEEIIWOAIDVLVOGKILLVYAGSSNPGYKIAOAMEIARRGTTIG 182
QY 183 MSHOCLYNLAHAELEVLPAQAAYGLGVAFMSPLHGLLSGALKEIAGTAVKSNQGR 242
Db 183 LVEOCLLYNLAERAEKEVYPAARDYGLGVIPMSPLHGLLSGVITKEA--TGGRRASGR 240
QY 243 AQVLL--PSLRPAIEAEKEFCRNIGEDPAEVLAWLSPRGIAVGIGPTPTPOLDLSALK 300
Db 241 AADALDKTKTREQIOAYVEDLIDKHGLDPCGEALAMLLTRPGVGTPIVGRPTGQDLSALR 300
QY 301 ASAMTDEQALSELDETFPAVAGGAPEAW 331
Db 301 ALELSEELSLDETFP--GPGSPPEAF 328

RESULT 7
O9PC04 PRELIMINARY: PRT: 329 AA.
ID O9PC04;
AC O9PC04;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SUGAR-PHOSPHATE DEHYDROGENASE.
GN XFI724.
OS Xylella fastidiosa.
OC Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9ASC;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Martins E.A.L., Madella H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
RA Vallada H., Van Sluys M.A., Verjowski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000)
DR EMBL: AE003996; AAF84533.1; -
DR InterPro: IPR001395; -
DR Pfam: PF00248; aldo_ket_red; 2.
SQ SEQUENCE 329 AA: 36339 MW: 6019CFEPD8037E1 CRC64;
```

```
QY 65 ---GTEELVGRMLAOGGREDTVALTKVGESESRVNDGSLARHIIASCEGSLRLIG 121
Db 63 KGRGTSEELYIGNMLADOKRRKROKIVLATKYQPMENGPDKTISAHIRACEASLKRK 122
QY 122 VDHIDVYQMHHDIRSAPWDEWQAMDSLVASGVSYSSNFRAGWHITAAQENARRHSL 181
Db 123 TDIHIDYQMHHDIDRAAPWDEWQAMDSLVASGVSYSSNFRAGWHITAAQENARRHSL 182
QY 182 GWYSHOCLYNLAHAELEVLPAQAAYGLGVAFMSPLHGLLSGALKEIAGTAVKSNQGR 241
Db 183 GLASEOCLYNLORTTELETPALRHFGIGLLPWSPTGMGLGVUKTASG---RRATP 239
QY 242 RAQVLLPSLRPAIEAEKEFCRNIGEDPAEVLAWLSPRGIAVGIGPTPTPOLDLSALK 301
Db 240 GLQAQIDKLRLPQLEAVEALCDELGQTPADVALAMLLHNVPVTAALSGPTVQDQNNLA 299
QY 302 SAMTDEQALSELDETFPAVAGGAPEAW 331
Db 300 LSVKLSDELTKLDEIWP--GPGGEAPQAY 327

RESULT 8
O9PC05 PRELIMINARY: PRT: 362 AA.
ID O9PC05;
AC O9PC05;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SUGAR-PHOSPHATE DEHYDROGENASE.
GN XFI723.
OS Xylella fastidiosa.
OC Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9ASC;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Martins E.A.L., Madella H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
RA Vallada H., Van Sluys M.A., Verjowski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000)
DR EMBL: AE003996; AAF84532.1; -
DR InterPro: IPR001395; -
DR Pfam: PF00248; aldo_ket_red; 1.
SQ SEQUENCE 362 AA: 39457 MW: DD224B68B59317D0 CRC64;
```



```

Query Match          45.8%; Score 791.5; DB 2; Length 362;
Best Local Similarity 48.6%; Pred. No. 1,2e-52;
Matches 162; Conservative 46; Mismatches 112; Indels 11; Gaps 4;

OY 5 AATHVRLGRSALLNSRLMGTVNFSGRVEDDALARLMDHARDRCINCLDTADMYGWM----60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 33 SGNFKHLGRFGLKYSRIALGCMNFGELTDEPNFRIMSALDSGVNLEPDTADYVGGPQT 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 61 RLYTGH--TEELVGRWLAAGGGRREDVYLAATKVGEMSERVNSGLSRHIIASCEGSLR 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 DMERGFTSEETIGRWLAQDRSRDRLLATKRYVOPMGTGNDYLLSAVHTRRACEASLTK 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 119 RLGDVHDVYGMHHIDNSAPMDEYQWAMDLSYASGVSYSSNFPAGHIIAAGENARR 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 RLKTDHIDLQMHVNDSTPQOELTQWAEQLVREGKITVYSSNFPAAADLARAQVASESR 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 179 HSLGMSHOCVLYNLAVHRAELEVLPAAQAVGLGVFANSPPLHGLLSGALKEKLAAGTAVKS 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 NPLGVSEQSLSYNIQRTIELEVIPAVRELIGLIPSPIMGLLGGVLTGKITEG---RR 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 239 AQAQAVLPSLRPAIAYEKEFCNLCGEDPAEVLAWLSRPGTAGAVIGRPTPEQLDSA 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 ATPGLQAOIQOFKRPDLAEYALCRELCOPPAVALAWVHNPTAAISGPRVEQOMREN 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 299 LKASAMTLDQALSELDEIFPAVASGGAPEAW 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 LKALSLMLSETLAKLDELTP--GPGGAPRAY 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
OY005408 PRELIMINARY; PRT; 316 AA.
AC 005408;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE YRPG.
GN YRPG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=168;
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., lauber J.,
RA Duesterhoeft A., Ehrlich S.D.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
MDLINE=98044033; Pubmed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Brann M., Bridnell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.O., Daniel R.A.,
RA Danilozot F., Devyane K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entlin K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Filiz S.Y., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nkal S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwolik S., Prescott A.M.,
RA Pesezan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
RA Sato T., Scanlan E., Schleich S., Schreuter R., Scoffone F.,

```

```

RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takematu K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler K., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U93875; AAB80886.1; -
DR EMBL: Z99117; CAB14626.1; -
SQ SEQUENCE 316 AA; 35475 MW; 207C59D9B0A9CED1 CRC64;

Query Match          44.28%; Score 763; DB 2; Length 316;
Best Local Similarity 50.6%; Pred. No. 1,4e-50;
Matches 161; Conservative 45; Mismatches 100; Indels 12; Gaps 6;

OY 19 SRLMLGTVNFSGRVEDDALARLMDHARDRCINCLDTADMYGWRLYKGTLELVGRWLAAG 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 SRLGLGMNFEVDVDEKAFRIMDEALDNGIQFEDTAIYGMGNAGTIESIIKWFAGOG 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 79 GGRREDVYLAATKVGEMSERVNS----GLSRHIIASCEGSLRRLGVDHIDVYGMHHI 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 GOREKVVYLAATKVEYPSD-PNDGPNDRKGLSLKIRRHLEGSLLKRLDTDIHELQOMHHI 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 134 DRSAPEWDEWQAMDLSYASGVSYSSNFPAGHIIAAGENARRHSLGMSHOCVLYNLA 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 DRRTPEIWEPAFTQVSGKVDYIGSNFAGMHLVKRAEAKRRFGLVTEQHKYSL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 194 VRHAEVLPAAQAVGLGVFANSPPLHGLLSGALKEKLAAGTAVNSAQRVOLLPSLRPA 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 ERTAEMEVLPAARDLGLGVVAMSPLAGLLGKALKSNAGT--RTAK-RAD-LIEKHRLQ 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 254 IEAVEKFCRNLCGEDPAEVLAWLSRPGTAGAVIGRPTPEQLDSALKASMTLDEQALSE 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 LEKFSDLCKELGEKEANVALAWLANVLTAPITIGPRVDEQRTIRAVELSDKEILRM 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 314 LDEIFPAVASGGAPEAW 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 LNDIFP--GPGGTERPEAY 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
OY09KXL8 PRELIMINARY; PRT; 284 AA.
AC 09KXL8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POTATIVE REDUCTASE.
GN SCD12A.22C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]

```



```

OY 179 HSLG-MVSHQCILYNNAVRAAELEVPAAQAYGLGVFAMSPILGLSGALEKTAGT---- 234
Db 174 HCLAFISMOSQSYNLIVRDEEDYVPLCEEBEIALTPMSPFLARGILLAGA---RSAGTLERT 230
OY 235 ----AVKSAOGRAOVLLPSLRPAIEAYEKFCRNIGEDPAEYGLAMVLSRPIAGAVIGP 289
Db 231 RTDEQAPRWYGGREY-----ESTIGALEKILAAARGPLPAQALAMWLLGRNGVAPIYGL 285
OY 290 RTPEQDSALRKSAMTLDDEQALSEID 315
Db 286 SRPHLEDAALALTLDLAEBCATLE 311

RESULT 13
OY 09L1S0 PRELIMINARY; PRT: 336 AA.
AC 09L1S0;
DT 01-OCT-2000 (TREMblrel, 15, Created)
DT 01-OCT-2000 (TREMblrel, 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel, 16, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCES6.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
CX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Brown S.P., Harris D.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Khasnshi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; ALJ38852; CAB72221.1; -.
DR InterPro; IPR001395; -.
DR Pfam; PF00248; aldo_ket_red; 2.
DR PRINTS; PR00069; ALDKETRDASE.
DR SEQUENCE 336 AA; 37060 MW; C68A60A31C9BB62C CRC64;

Query Match 24.8%; Score 428.5; DB 2; Length 336;
Best Local Similarity 34.0%; Pred. No. 5.5e-25;
Matches 117; Conservative 48; Mismatches 140; Indels 39; Gaps 7;
OY 8 HVRLGRSALLTSRLWLGTVNFS-----GRVEDDALRLMDHARDGINCIDTAD 56
Db 3 YIRLGTGTELEVSALIALGCMGCGEPDRCGEPMISLGADPSRDITR---QALEGGVNFIDTAN 59
OY 57 MYGMWLYGNHIEELVGNRLAOCGGRRREDTVALATKYGGEMSEKSEVNDSCLSARIITASCES 116
Db 60 GYS----AGNSEETIGQ-AVRDFAEREEVYVSTKWMRRRPGPNAGLSRKAIPELDS 114
OY 117 LRLGVCHDIIDYOMHIDRSAPWDEWQAMDSLVASGRVSYGSSNPFNGMHTAAQENNA 176
Db 115 LKRIGTDYIIDLYQIHRMDYDPIETTELPHLDYVASCGRVYRIGASSWYAWQFAKALYLD 174
OY 177 RRHSLGMYSHOCLYMLAVRHALELEVLPAAQAYGLGVFAMSPILHGLSGLALEKTAGTAV 236
Db 175 LMGWTFPVSQMDHYVLIHREAREKMLPICADQDGVIVFPMSPFLARGILRRAR---TATAR 231
OY 237 KSAOGRAOVLLPSLRPAI-EAYEKFCRNIGEDPAEYGLAMVLSRPIAGAVIGPTPEOL 295

```

```

Db      232 AGTDEGGGATILYDEDDQAAVAERVRREIAGRGRSLSPAQVALAMVWRNRPVPSIVGYTKPAQL 291
Oy      296 DSALKASAMTIDEOALSLEDF-----PANAS 323
Db      292 ADAIAAVDGLDEDAALTEEPYOPHEAAYLEESFYKARPAAS 335

RESULT 14
O9RUC4 PRELIMINARY; PRT: 307 AA.
AC O9RUC4
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
PT PUTATIVE ALDO/KETO REDUCTASE (FRAGMENT).
GN SCF81_28C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_taxid=1902;
LN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denaplatte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL133171; CAB61562.2; -.
DR InterPro; IPR001395; -.
DR InterPro; IPR001822; -.
PF pfam: PF00248; aldo_ket_red: 1.
PRINTS: PR00069; ALDKERTDRAFTSE.
DR PROSITE; PS00398; RECOMBINASES_2; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 307 AA; 34107 MW; EC2D2594595FB18 CRC64;

Query Match 24.4%; Score 421; DB 2; Length 307;
Best Local Similarity 36.0%; Pred. No. 1.8e-24;
Matches 102; Conservative 46; Mismatches 127; Indels 8; Gaps 5;
Oy      34 DDDALR-LMDHARDGINCLDPTADMYGWRLKGHTGHTEELVGRMLAAGGGRRREDTVLATRGV 92
Db      21 DEASRPRLIRQALEGVGFFPDANYNS----DGTSBELYGRALAD-FARRDDIVLATRVN 75
Oy      93 GEMSERVNDGSLSANRHITASCCEGSLRLRGVDHIIDYQNHMHIDRSAPMDEWQAMDLSVAS 152
Db      76 GRMRPGPAGAGISRRRAVTEIEDHSIRRLGTVDYLYQHREDPPHPVEETMALHDLVKA 135
Oy      153 GKVSIVGSSNFAGMIIAAOENAAHRSHSLGMWSHOCTNLNAVRYNAELEVLPAQOVYGCV 212
Db      136 GKVRIGASSMATAMOFCKMOTTAERHGHTKFKYSMONHNINLYIREEREKMLPCADQGVG 195
Oy      213 FAWSELHGLLSGALEKLIAAGTAVKSAQGRAOVLPLSRPALAEAKFCRMNGEDPAEVG 272
Db      196 LPMSPLAGRILTRDM-D-TATGRSATDTEFG-SPLYEGCBRAVYEAVTRIAGERGVPRARA 253
Oy      273 LAMVLSRCGIAGAVIGRTPPEDLDALKASAMTIDEOALSLED 315
Db      254 LAMLLHQDTVAPIVGA SRPHGLEDAVAVALVELTSETIEEILE 296

```

```
RESULT 15
09KU57 PRELIMINARY: PRT: 352 AA.
AC 09KU57:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OXIDOREDUCTASE TAS, ALDO/KETO REDUCTASE FAMILY.
GN VC0667.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_taxid=666;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE-20406833; PubMed-10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004153; AAF93832.1; -
DR TIGR: VC0667; -
DR InterPro: IPR001395; -
DR Pfam: PF00248; aldo_ket_red: 1.
SQ SEQUENCE 352 AA; 39549 MW; 9E37E3F477BAEA98 CRC64;
```

Query Match 24.0%; Score 414.5; DB 2; Length 352;  
Best Local Similarity 34.4%; Pred. No. 6.8e-24;  
Matches 116; Conservative 56; Mismatches 140; Indels 25; Gaps 7;

```
OY 5 AATHVRLGSRALLTSTRIMLTGVNFGSRVEDDALLRLMDHARDGINCLDTADMY---GWR 61
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 8 AMQYTKLPHSSLEISKICIGTWTFFGQNSQADAFQOLDVALERGVTITAEKTPVPTA 67
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 62 LYKHTTELVGFWLAOGGGRREDTVLATKVGEMS-ERYNDS-GLSARHIIASCEGSLRR 119
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 68 OTGKTEEFITGMWLAK-SGKREKIYATKAGPRNVPIYIDKMDLDRNIIHQAVDDSLR 126
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 120 LGVDHIDVQMHIDRSA-----PMD-----EVMQAMDSIVASGKVSIVGSSN 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 127 LQTDYIDLYQLHWPQRQNTFGOLNVPYDPKQEEVTLLETLALNDLVPMGKVRRTIGVSN 186
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 163 FAGMHIIAAQAENARRHSLGMYSHOCLVNLAVRHALEVLPAQAQVGLGFAMSPPLHGCL 222
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 167 ETPMGVMSYLRLEKHELPRIYVSIONPYLLNRSEFVGLAEISHLGVLKLAYSPLAFGA 246
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 223 LSGALEKLA--AGTAVKSAQGAQVLLPSLRPAIEAYEKFCRNIGEDPAEVLAWVLSRP 280
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 LSGKYLNGARFAGARCTLHQRFSSRYFTEGILATAYVALAQOFGHDPQAMALAFVNGRP 306
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 281 GIAGAVIGPRTPPEQDLSALKASAMTLDQALSELDEI 317
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 307 FVASNITGATTMEQLKSLNDSLISLNEELLOKIOEI 343
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: June 22, 2001, 09:25:45  
Job time: 110 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:24:10 ; Search time 11.8 Seconds  
(without alignments)  
966,701 Million cell updates/sec

Title: us-09-463-705a-2  
1728

Sequence: 1 MTTDATHVRLGRSALTISR.....LDEIFPAVASGGAPEAWIQ 333

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	24.2	324	1	YAOJ_ECOLI
2	334	19.3	346	1	TAS_ECOLI
3	312	18.1	387	1	STCV_EMENT
4	307	17.6	307	1	A115_TOBAC
5	303.5	17.6	310	1	YCCB_BACU
6	293.5	17.0	331	1	G569_BACU
7	286	16.6	310	1	IOIS_BACU
8	285	16.5	306	1	IN22_MAIZE
9	285	16.5	388	1	NORA_ASPEL
10	283	16.4	388	1	NORA_ASPEL
11	273	15.8	376	1	AADE_YEAST
12	268	15.5	385	1	AADE_YEAST
13	263.5	15.2	329	1	AAAD_YEAST
14	256.5	14.8	340	1	YAKC_SCHPO
15	255.5	14.8	306	1	YOKF_BACU
16	255.5	14.8	363	1	AAAD_YEAST
17	253.5	14.7	323	1	Y98_MYCTU
18	244.5	14.1	300	1	YCSN_BACU
19	244.5	14.1	326	1	YDUG_ECOLI
20	217.5	12.6	330	1	AR72_HUMAN
21	203.5	11.8	285	1	ALKE_RABBO
22	202.5	11.7	331	1	AR72_HUMAN
23	198	11.5	288	1	AAAD_YEAST
24	196.5	11.4	327	1	AAAD_YEAST
25	184	10.6	267	1	YAFB_ECOLI
26	169	9.8	298	1	YDHF_ECOLI
27	168.5	9.6	286	1	YDHC_ECOLI
28	162	9.4	284	1	P100_LEIMA
29	161.5	9.3	318	1	XYLI_PICST
30	158	9.1	275	1	YDHE_ECOLI
31	157	8.9	277	1	2DKG_CORSP
32	153.5	8.9	324	1	ALDX_RAT
33	153.5	8.8	312	1	YPR1_YEAST

34	152.5	8.8	324	1	ALDX_PIG	P50578	sus scrofa
35	150.5	8.7	295	1	MORA_PSEPU	O02198	pseudomonas
36	149.5	8.7	324	1	ALDX_HUMAN	P14550	homo sapien
37	149	8.6	310	1	S6PD_MALDO	P28475	malus domes
38	148.5	8.6	320	1	ALDR_HORVU	P23901	hordeum vul
39	147	8.5	321	1	YDGT_SCHPO	O10494	schizosacch
40	144.5	8.4	312	1	GCY_YEAST	P14065	saccharomyc
41	132.5	7.7	282	1	Y166_YEAST	P47137	saccharomyc
42	129.5	7.5	103	1	YCR4_ERWHE	O01333	erwinia her
43	127	7.3	318	1	XYLI_PACRA	P78736	pachysolen
44	125.5	7.3	329	1	XYLI_KLUVA	P43378	kluveromyc
45	125	7.2	212	1	AADE_YEAST	P43547	saccharomyc

## ALIGNMENTS

RESULT 1	YAOJ_ECOLI	STANDARD:	PRT:	324 AA.
ID	YAOJ_ECOLI			
AC	P77735:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	HYPOTHETICAL OXIDOREDUCTASE IN PGPA-1SPA INTERGENIC REGION.			
GN	YAOJ.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATN-K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12";			
RL	Science 277:1453-1474(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,			
RA	Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurd O.,			
RA	Lew H., Lin D., Namath A., Oether P., Schramm S., Davis R.W.;			
RL	Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.			
CC	- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; AE000148; AAC73522.1; ALT_INIT.			
DR	EMBL; U82664; AAB40175.1; ALT_INIT.			
DR	Ecogene; EG13611; YAOJ.			
KW	Hypothetical protein; Oxidoreductase.			
FT	ACT SITE 127 HYDROGEN-BOND DONOR (PROBABLE).			
SO	SEQUENCE 324 AA; 36420 MW; 0869E09F2BED9B1 CRC64;			
Query Match	24.2%; Score 418; DB 1; Length 324;			
Best Local Similarity	34.6%; Pred. No. 8e-26;			
Matches 112; Conservative 52; Mismatches 124; Indels 36; Gaps 7;				
OY	11 LGRSALTSRLVLTGVNF-----SGRVEDDRLRLMDHARDGINCLOTADYGMWL 62			
DB	6 LKKTDLRVSRLCLGCMFCEPDGRGNHAWTLPPRESSRPIIKRALEGINFEDTANSYS--- 62			
OY	63 YGHTHEELVGRRLAAGCGGRREDTIVLATKVGGMSEKRVND--SGLSARHTIASCEGSLRL 120			

```

Db      63 -DGSSEIQRAL-RDFARREDVAVTKV-----FHRVGDIPEGLSRAQLIRSIDSLRL 116
      121 GVDHIDVYOMHIDR-APDEWQAMDSLVASKSYVSSNAGNHIAAOENARRHS 180
      117 GMDYDIIDIHREDYTPLEETLEALNDVYKAGKARYIGASSMHSQFOALELOROHG 176
      181 LGVSHQCYLNLAVRAAELEVLPAQAQYLGAVFAMSPHGLLSG-----ALEKLA 231
      177 AGFVSNQMDHYNLIRREEREMLPICVOEGAVIIPMSPLARGRLTRRWGETTARLVSD 236
      232 AGTAVKSAOQAVLLPSLRPALEAYEKFCRNIGEDPAEYGLAWLSPGICAGAVIGPRT 291
      237 KNLKRESDEMDAQI-----AERLTGVSEELGATRAQVALMLLSKPGIAPITGTSR 288
      292 PEQDLSAKASAMTLDEQALSELD 315
      289 EROJDELINAVDITTLKPEQIAELE 312

```

## RESULT 2

TAS\_ECOLI  
ID TAS\_ECOLI STANDARD: PRT: 346 AA.

```

AC 046833;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE TAS PROTEIN.
GN TAS.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=562;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis J., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [12]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-B/R / W03610;
RX MEDLINE-98240727; PubMed-9560382;
RA Timms A.R., Bridges B.A.;
RT "Reversion of the tyrosine ochre strain Escherichia coli W03610 under
RT starvation conditions depends on a new gene tas."
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: 029561; AAB0481.1;
DR EMBL: AE000367; AAC75873.1;
DR EMBL: Y14609; CAA74961.1;
DR HSSP: P14550; ZALR.
DR Ecocore: KGI3093; tas.
DR InterPro: IPR001395;
DR Pfam: PF00248; aldo_ket_red; 1.
FT Oxioreductase.
KW ACT SITE 132
SQ SEQUENCE 346 AA; 38499 MW; 111692D06CA07CD7 CRC64;

```

Query Match 19.3%; Score 334; DB 1; Length 346;  
Best Local Similarity 30.4%; Pred. No. 3.4e-19;  
Matches 106; Conservative 62; Mismatches 135; Indels 46; Gaps 10;

```

QY 10 RIGRSALLTSLRLGTVNFSGRVEDDALRLMDHARDGICLDTADMY----GWRLYKHG 66
      5 RPHSSLEVVSTLGLTGMTGECNSEADAHADLYAAAGINILIDVEMVPPRPETOGI 64
      67 TEELVGRMIAOAGGRRDPTVLATKVGEMSERVNDG-----LSRHIIASCEGSLRL 120
      65 TETVGNWMLAKHGS-REKLIISKVSG--PSRNDGCIIPDQALDKRNIREALHDSLKRL 121
      121 GVDHIDVYOMHIDR-----SAPWDEWQAMDSLV--ASGKSYVSSSNF 163
      122 QTDYLDLYGVHWPQRPNTCEGKLGYSMTDSAPAVSLDTLDALAEYORAGKIRYIGVSN 181
      164 AGNHIAAAGENARRHSLGNSHQCYLNLAVRAAELEVLPAQAQYLGAVFAMSPHGLL 223
      182 TAFGVMRVYHLADKHDLPRVTIQNPYSLNRSFEGVLAESQYEGVELLAISGLGFGTL 241
      224 SGALERKLAAGTAVKSAOQAVLLPSLR-----PALEAYEKFCRNIGEDPAEYGLAW 276
      242 TG---KYLNG--AKPAGARNTLFSRFTRSGEQTKAQAAYAVDIARRHGDPRQMALAV 296
      QY 277 LSRPGIAGAVIGPRTPEQDLSAKASAMTLDEQALSELDI-----FPA 320
      297 RROPFVASTLLGATMTDQKTNIESLHLESEEDVLAIEAVHGYVTPA 345

```

## RESULT 3

STCV\_EMENT  
ID STCV\_EMENT STANDARD: PRT: 387 AA.

```

AC 000727;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV
DE (EC 1.1.1.-).
GN STCV.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
ON NCBI_TaxID=5072;
RX MEDLINE-96202293; PubMed-8643646;
RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA Keller N.P., Adams T.H., Leonard T.J.;
RT "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT cluster in Aspergillus nidulans."
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).

```

```

CC -1- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF 5-
CC HYDROXYAVERAMIN TO FORM AVERPIN.
CC -1- PATHWAY: STERIGMATOCYSTIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: U34740; AAC49206.1;
KW Oxioreductase.
FT ACT SITE 148
SQ SEQUENCE 387 AA; 43548 MW; 1621588273B85E8C CRC64;

```

Query Match 18.1%; Score 312; DB 1; Length 387;  
Best Local Similarity 30.2%; Pred. No. 2.1e-17;

Matches 98: Conservative 57: Mismatches 135: Indels 34: Gaps 10:

OY 14 SALLTSRLMGTVNFSR-----VEDDRLMDHARDGICLDTADMGWLYYGH 66  
 DB 23 AGIRVSPCLGTMHFGCGMMRAMGVDVTKETAFALLDREYEAAGNFIDTAN----FYQGE 77  
 OY 67 -TEEVYGRMLACGGGRRDVTATK-----VGGEMSERVNDGSLARHIIASCESGLR 119  
 DB 78 GSEKMLGEMVA-SRGNRDELATATYTMSTYLTGPEKIKSNQSGSHSLRSLVASTAK 136  
 OY 120 LGVDHIDVYOMHHIDRSAPMDEVMQAMDSLVAAGKVSYSNFGMWHIAAOENARRH 179  
 DB 137 LATDYIDLLYHMPDFSTSVEEVMQSLHHLVAAGKVLNIGISDAPAMVAKNE-YARRH 195  
 OY 180 SLGMYSHOCLYN---LAVRAHEVLPAQAYGLGVFAMSPFHGGLSLGALEKLAAGTA 235  
 DB 196 GL---TRCVYOGRWACSYRDFEREILPMCOSEGLALAPWGLGROYKSAEEFOEGTR 252  
 OY 236 VKSAGRAQVLLPSLRPAIEAEKFCRNLDGEPAEVLAVLWVRGAGAVGPTPEQL 295  
 DB 253 NMGPQEEKHRLMGA--KLEVEGER---KGVAALALATLHKSPPYFVPIGCRTEQL 306  
 OY 296 DSALKASAMTDEQALSELDEIFP 319  
 DB 307 EANITSLGVELSDEIYEIEDTIP 330

## RESULT 4

AL15\_TOBAC STANDARD: PRT: 307 AA.

AC P40691:  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE AUXIN-INDUCED PROTEIN PCNT115.  
 OS Nicotiana tabacum (Common tobacco)  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV WHITE BURLLEY;  
 RA van der Zaai E.J.;  
 RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.  
 CC CC  
 CC -1- INDUCTION: BY AUXIN.  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; X56267; CAA39708.1; -  
 DR PIR; S16390.  
 DR InterPro: IPR001395; -  
 DR Pfam; PF00248; aldo\_ket\_red; 1.  
 KW Oxidoreductase.  
 FT ACT\_SITE 136  
 FT SEQUENCE 307 AA; 33857 MW; C4417DA852613360 CRC64;

Query Match 17.8%; Score 307; DB 1: Length 307;  
 Best Local Similarity 30.1%; Pred. No. 3.9e-17;  
 Matches 92: Conservative 50: Mismatches 136: Indels 28: Gaps 8;

OY 9 VILGRSALLTSRLMGTVNFSG---RVEDDRLMDHARDGICLDTADMGWLYYK 64  
 DB 11 IRLSGGLEVAHQGLGCMGMSAFYGPAPKPEPMIQLIHAIINSGLTLDTSVYG----- 65

OY 65 GHTEE-LVGRMLAOGGRRDVTATKVGEMSERVNDGSLARH-----ITASCEGSLR 118  
 DB 66 PHTNELLLGK--ALGGTRERVALTKFGIYLGDEKAKGKAAVGDPAVYAAACEASLK 123  
 OY 119 RGVHDIDVYOMHHIDRSAPMDEVMQAMDSLVAAGKVSYSNFGMWHIAAOENARRH 178  
 DB 124 RLDIDICIDLYHRDTRPFIETVYGLKKLYBECKLAKTYGSE-----ASASTIRRA 176  
 OY 179 HSLGAVSH-QCLYNLAVRAHEVLPAQAYGLGVFAMSPFHGGLSLGALEKLAAGTAVK 237  
 DB 177 HAVHPIITAVQLWMSLMSRDVEEIIPTCRELGIGIYAYSPLRGFLSSGPKLEDMNSND 236  
 OY 238 SAQGRAQVLLPSLRPAIEAEKFCR---NLGSDPAEVLAVLWVRGAGAVGPTPEQL 294  
 DB 237 YKKYLPFRQAELENKKNLYERICEAVARKGCTPSOLALAWHOGQNDVCPITGTTKEN 296  
 OY 295 LDSALK 300  
 DB 297 LNOXMK 302

## RESULT 5

YCKK\_BACSU STANDARD: PRT: 310 AA.

AC P46905:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHETICAL OXIDOREDUCTASE IN NATB-RAPD INTERGENIC REGION (ORF).  
 GN YCKK.  
 OS Bacillus subtilis.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunano M., Tamakoshi A., Yamane K.;  
 RT "A 32 kb nucleotide sequence from the region of the lincomycin-  
 RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and  
 RT identification of the site of the lin-2 mutation."  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-268 FROM N.A.  
 RC STRAIN=BD99;  
 RA Cheng J., Guffanti A.A., Krulwich T.A.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC CC  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AB000617; BAA22238.1; -  
 DR EMBL; Z99105; CAB12071.1; -  
 DR EMBL; U30873; AAB53024.1; -  
 DR HSSP; P80276; IAH3.  
 DR Subtilist; BG11340; YCKK.  
 DR InterPro: IPR001395; -  
 DR Pfam; PF00248; aldo\_ket\_red; 1.  
 KW Hypothetical protein; Oxidoreductase.  
 FT ACT\_SITE 127  
 FT SEQUENCE 310 AA; 34800 MW; C06BF4195D235C91C CRC64;

Query Match 17.6%; Score 303.5; DB 1: Length 310;  
 Best Local Similarity 28.9%; Pred. No. 7.4e-17;  
 Matches 92: Conservative 54: Mismatches 147: Indels 25: Gaps 8;

[illegible]

RESULT	6
GS69_BACSU	STANDARD: PRT: 331 AA.
ID	GS69_BACSU
AC	P80874; 007583;
DT	01-NOV-1997 (Rel. 35, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	15-JUL-1998 (Rel. 36, Last annotation update)
DE	GENERAL STRESS PROTEIN 69 (GSP69).
GN	YHDN
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1423;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RA	NOACK M.A., Terpstra P., Holtsappel S., Venema G., Bron S.;
RL	Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
RN	[2]
RP	SEQUENCE OF 1-25.
RC	STRAIN=1558;
RA	MEDLINE=97443988; PubMed=9298659;
RA	Antelmann H., Bernhard J., Schmid R., Mach H., Voelker U.,
RT	Hecker M.;
RT	"First steps from a two-dimensional protein index towards a response-
RL	regulation map for Bacillus subtilis.";
CC	Electrophoresis 18:1451-1463(1997).
CC	-I- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
CC	LIMITATION AND OXYGEN LIMITATION.
CC	-I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement. See <a href="http://www.isb-sib.ch/announce/sib-sib.ch">http://www.isb-sib.ch/announce/sib-sib.ch</a> .
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .
CC	-----
DR	EMBL: Y14082; CAA74498.1; -
DR	EMBL: 299109; CAB12792.1; -
DR	HSSP: P23457; LMT.
DR	Subtilist; BG13020; yhdN.
DR	Pfam; PF00248; aldo_ket_red; 1.
KW	Oxidoreductase; Heat shock.
FT	ACT_SITE 125 125
FT	CONFLICT 25 25 HYDROGEN-BOND DONOR (PROBABLE).
FT	SEQUENCE 331 AA; 37312 MW; 82BC24D46E4994D CRC64;
Q0	VSUENICE 25 25 G -> K (IN REF. 2).

```

Query Match Similarity 17.0%; Score 293.5; DB 1; Length 331;
Best Local Similarity 26.2%; Pred. No. 4,9e-16;
Matches 88; Conservative 67; Mismatches 150; Indels 31; Gaps
8

QY      8 HVRLGRSALLTSRLMLGVNFSGRV----EDDALARLMDHARDRGINCITDADMYQWRDY 63
DB      3 YTSINDTGEIARIGIGTGAIGTWMGCTGDEKTSIETIHAALDQGITLIDTPAVGFG--- 59

QY      64 KHTTELVGKRLAOGGGRRRDTVLATKVGGEWSERYNDGSLSRHITLACBGLKLVSD 123
DB      60 -GQSEIYGAKE-KYGRKDOYLATKTLDMKNNOLFPHANRARIIVEVENSLSKLQTD 117

QY      124 HLDVYOMHHIDSSAPMDEWQAMDSIVASGKSYVSSGAFGHIAAENAAHRSLDM 183
DB      118 YIDLQVMPHDDPLVPLEETAAYMKELYDQKTRATGVSNF-----STEDMFRVAVPL 171

QY      184 VSHOCLYNLAVNHAIELEVPAQAAYGLGVAFWSPHLHGGLSGL--EKLAAGTAVKSAQG 241
DB      172 HTIQPPYNNLEFREMESVLPYAKDKKITLLXGSLDRGLLTGKMEETLYFFEGDDLNRHP 231

QY      242 RAQVLLPSLRPAIEAV-----EKFCR- NLGEDPAEVLGLAWLSPRGITAGAVIGRTPE 293
DB      233 KFG-----KPRFEKELSAVNOLDKLAKRYRKSVINHLAVRWIIDDPGADIALMGARKPG 285

QY      294 QLDSALKASAMTLLDEQALSELDEIFPAVAGGAPE 329
DB      286 QLEALSETIGTWLNSEDOKDINTILENTISDPVGE 321

```

```

RESULT 7
ID IOLS_BACSU STANDARD: PRT: 310 AA.
AC P46336;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE IOLS PROTEIN (VEGETATIVE PROTEIN 147) (VEG147).
GN IOLS OR SS92ER.
OS Bacillus subtilis.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / BGSC141;
RX MEDLINE=96093926; PubMed=7584049;
RA Yoshida K.-I., Seki S., Fujimura M., Miwa Y.,
RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RT genome between the gnt and iol operons.";
RL DNA Res. 2:61-69(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / BGSC141;
RA Fujita Y., Shibayama T., Ishio I., Aoyama D., Yoshida K.-I.;
RT "Organization and transcription of the myo-inositol operon, iol, of
RL Bacillus subtilis.";
RN Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-20.
RC STRAIN-IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -I- PATHWAY: MYO-INOSITOL CATABOLISM.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

```



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AB005554; BAA21607.1; -  
 DR EMBL; 299124; CAB16014.1; -  
 DR HSSP; P23457; LIMI.  
 DR Subtilist; BG11363; 101S.  
 DR Pfam; PF00248; aldo\_ket\_red; 1.  
 DR Oxidoreductase.  
 FT ACT\_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).  
 SO SEQUENCE 310 AA; 35168 MW; A870F226F8684867 CRC64;

Query Match 16.6%; Score 286; DB 1; Length 310;  
 Best Local Similarity 28.6%; Pred. No. 1.8e-15;  
 Matches 95; Conservative 58; Mismatches 129; Indels 50; Gaps 11;

QY 10 RLGSALLTSLMTGTVNFSG-----RVEDDALRLMDHARDRGINCIDTADMTGWLK 64  
 Db 5 KLGSDDLQVFPFGLGTNAVGSHNLYPNLNETGKELVREAIRNGVTMDTAYIG-----I 60  
 QY 65 GHTSELVGRMLAOGGREDVTATKVG--GEMSERVNDGSLSAHITASCBSLRIGV 122  
 Db 61 GRSELLEGEVIRE--FNEDEVIAATKAHRRKQGNDEVDN--SPDELFKSVDESILRLNT 116  
 QY 123 DHIVYQMHIDRSAPWDEWQAMDSLVAAGKVSYSVSSNFGWHITAAQENARRHSIG 182  
 Db 117 DYIDLFYHPEDEHPKQEAVALNEMKAKKIRSIGVSNFSLDLKRNAND-----G 169  
 QY 183 WVS-HQCLYNLAVRHALEVLPAQAQGLGVFAMSPHGLGSLGALAKTAVKASAG 241  
 Db 170 LVVDLQGEYNNILNRAEKTFFPYTKENHISPIPYFPLVSGLAG--KYTEDTPEPS-- 224  
 QY 242 RAQVLLPRLRAIEAY--EKFCRNIGE-----DPAEGLAWVLSRQIAGV 286  
 Db 225 -----DLRNEQEHFGEKERENKRNKLAFLPAEKHNVDIPHVLAWYLARPEIDILI 277  
 QY 287 IGPRPEGLDSALKASAMTLDQALSELDEIF 318  
 Db 278 PGAKRADQLINIKTADVTLSQEDISFIDKLF 309

RESULT 8  
 IN2\_MAIZE STANDARD; PRT; 306 AA.  
 AC P49249;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE IN2-2 PROTEIN.  
 GN IN2-2 OR SAF2.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;  
 OC Andropogoneae; Zea.  
 OC NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. MISSOURI 17;  
 RX MEDLINE=92003682; Pubmed=1912492;  
 RA Hershey H.P., Stoner T.D.;  
 RT "Isolation and characterization of cDNA clones for RNA species  
 RT induced by substituted benzene sulfonamides in corn."  
 RL Plant Mol. Biol. 17:679-690(1991).  
 CC -1- TISSUE SPECIFICITY: LEAF AND ROOT.  
 CC -1- DEVELOPMENTAL STAGE: IT APPEARS IN ROOTS WITHIN 30 MIN OF  
 CC INDUCTION, MAXIMUM LEVELS ARE REACHED BY 6 HRS, AND REMAINS  
 CC CONSTANT FOR 2 DAYS. IN LEAVES IT IS SEEN 9 HRS AFTER INDUCTION,  
 CC AND REACHES MAXIMUM LEVELS AFTER 24 HRS.  
 CC -1- INDUCTION: BY N-(AMINOCARBONYL)-2-CHLORO BENZENE SULFONAMIDE (2-

CC (CBSU).  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.  
 DR MaizeDB; 121988; -  
 DR Oxidoreductase.  
 FT ACT\_SITE 131 131 HYDROGEN-BOND DONOR (PROBABLE).  
 SO SEQUENCE 306 AA; 33828 MW; 2E94674CFDE3CDB CRC64;

Query Match 16.5%; Score 285; DB 1; Length 306;  
 Best Local Similarity 32.9%; Pred. No. 2.1e-15;  
 Matches 73; Conservative 40; Mismatches 89; Indels 20; Gaps 6;

QY 9 VRLGSALLTSLMTGTVNFSG-----RVEDDALRLMDHARDRGINCIDTADMTGWLK 64  
 Db 11 KLGSQGLEVSAQGGCGMGSFAFYGPPESEMKRLHNAVDAGVTFTDSDVYG----- 65  
 QY 65 GHTSELVGRMLAOGGREDVTATKVGEMSERVNDGSLSAHITASCBSLRIGV 123  
 Db 66 PHTNVLVIGK--ALOGGVREKVELATKFGVSFADGKREIHGDPAVYTRACBSFRRLGVD 123  
 QY 124 HIDVYQMHIDRSAPWDEWQAMDSLVAAGKVSYSVSSNFGWHITAAQENARRHSIG 183  
 Db 124 CIDLYQHRIDKRVPIETVIGELKLVBEGRKIKYIGLSE-----ASASTIRRAHAYHP 176  
 QY 184 VSH-QCLYNLAVRHALEVLPAQAQGLGVFAMSPHGLGSL 224  
 Db 177 ITAVOLEWMSLRDAEEDIIPTCRLGIGIYAVSPLGGRFSS 218

RESULT 9  
 NORA.ASPFL STANDARD; PRT; 388 AA.  
 AC Q00049;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NORSOLORINIC ACID REDUCTASE (EC 1.1.1.1-).  
 GN NORA OR ADH-2.  
 OS Aspergillus flavus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OC NCBI\_TaxID=5059;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NREL 3357;  
 RX MEDLINE=96156784; Pubmed=8593042;  
 RA Cary J.W., Wright M., Bhatnagar D., Lee R., Chu F.;  
 RT "Molecular characterization of an Aspergillus parasiticus  
 RT dehydrogenase gene, nora, located on the aflatoxin biosynthesis gene  
 RT cluster."  
 RL Appl. Environ. Microbiol. 62:360-366(1996).  
 CC -1- PATHWAY: AFLATOXIN BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; U32377; AAC49167.1; -  
 DR Oxidoreductase.  
 FT ACT\_SITE 148 148 HYDROGEN-BOND DONOR (POTENTIAL).  
 SO SEQUENCE 388 AA; 43751 MW; 758B8187187BF580 CRC64;

Query Match 16.5%; Score 285; DB 1; Length 388;  
 Best Local Similarity 28.6%; Pred. No. 2.8e-15;  
 Matches 97; Conservative 58; Mismatches 132; Indels 52; Gaps 12;

QY 8 HVRLGRSA-LTSLMTGTVNFS-----GRVEDDALRLMDHARDRGINCIDTADMTG 59

```

Db      16 HLLSSASAVRSPCLCLGTMSEFGKGVMGCEDOATSFENMLDTFEYSGNFIDVANYF- 74
OY      60 WRLYKHTHELINQWRLAOGGGRREDIVLATK-----VGGEMSERVNDGSLSRHIIASG 113
Db      75 ---OGGDTERWGWEMMAQ-RQNDEIVLSTKYTMGPQKIRKSNFQGNHAKSLRLSV 130
OY      114 EGSRLRLGVHDIDVYOMHIDRSAPWDEWQAMDSLVAASKVSYGSSNFAGMHIAAOE 173
Db      131 KASLQKLOTIDYIDLVLVHMDDFTTSVEEYMRSLNHLVANGKVLVLYGSDTPALVYKC-- 188
OY      174 NA-ARRHSLGWS-HOCLYNLAVRHALEVLPAQAYGLGVFAMSPDLHGSLGALKEKLA 231
Db      189 NAFARANGLTPEFSVYOGHWSAFRDEPDLPMCESESGMLAPWGLVGRQFRSAEE--- 245
OY      232 AGTAVASAGRAQVLLPSLRPAIEAEKFCRNIGE-----DPAEVLAWLNSRP 280
Db      246 -----FSREGR-----KMGPODEKH-----RLGEKLDQMAOQKNTKATSIQAYVMHKA 290
OY      281 GIAGAVIGRPTPEQDLSALKASAMTLEQALSELDEIFP 319
Db      291 PYEPIYIGRKVEHLKENIEALGLVLSSEIRIIDAEP 329

```

```

RESULT 10
NORA_ASPPA STANDARD; PRT; 388 AA.
AC 000258;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE NORSOLORINIC ACID REDUCTASE (EC 1.1.1.-).
GN NORA.
OS Aspergillus parasiticus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 163 / NRRL 5862 / SU-7;
RX MEDLINE=96156784; PubMed=8593042;
RA Cary J.W., Wright M., Bhattacharjee D., Lee R., Chu F.;
RT "Molecular characterization of an Aspergillus parasiticus
RT dehydrogenase gene, nora, located on the aflatoxin biosynthesis gene
RT cluster."
RL Appl. Environ. Microbiol. 62:360-366(1996).
CC -1- PATHWAY: AFLATOXIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U24698; AAC49166.1; -.
KW Oxidoreductase.
FT ACT_SITE 148
SQ SEQUENCE 388 AA; 43721 MW; 8829491E66EF6BB CRC64;

```

Query Match 16.4%; Score 283; DB 1; Length 388;  
 Best Local Similarity 28.6%; Pred. No. 4e-15;  
 Matches 97; Conservative 58; Mismatches 132; Indels 52; Gaps 12;

```

OY      8 HVLGSA-LTSLRMIGVNFPS-----GRVEDDALRLMDHARDGNCIDTDMWG 59
Db      16 HRLSSAGVRSPLCLGTMSEFGKGVMGCEDOATSFENMLDTFEYSGNFIDVANYF- 74
OY      60 WRLYKHTHELINQWRLAOGGGRREDIVLATK-----VGGEMSERVNDGSLSRHIIASG 113

```

```

Db      75 ---OGGDTERWGWEMMAQ-RQNDEIVLSTKYTMGPQKIRKSNFQGNHAKSLRLSV 130
OY      114 EGSRLRLGVHDIDVYOMHIDRSAPWDEWQAMDSLVAASKVSYGSSNFAGMHIAAOE 173
Db      131 KASLQKLOTIDYIDLVLVHMDDFTTSVEEYMRSLNHLVANGKVLVLYGSDTPALVYKC-- 188
OY      174 NA-ARRHSLGWS-HOCLYNLAVRHALEVLPAQAYGLGVFAMSPDLHGSLGALKEKLA 231
Db      189 NAFARANGLTPEFSVYOGHWSAFRDEPDLPMCESESGMLAPWGLVGRQFRSAEE--- 245
OY      232 AGTAVASAGRAQVLLPSLRPAIEAEKFCRNIGE-----DPAEVLAWLNSRP 280
Db      246 -----FSREGR-----KMGPODEKH-----RLGEKLDQMAOQKNTKATSIQAYVMHKA 290
OY      281 GIAGAVIGRPTPEQDLSALKASAMTLEQALSELDEIFP 319
Db      291 PYEPIYIGRKVEHLKENIEALGLVLSSEIRIIDAEP 329

```

```

RESULT 11
AADE YEAST STANDARD; PRT; 376 AA.
AC P42884;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE ARYL-ALCOHOL DEHYDROGENASE AAD14 (EC 1.1.1.-).
GN AAD14 OR YNL331C OR N0300.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA Van Dyck L., Pascual-Ahuir A., Goffeau A.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY. STRONG,
CC TO YEAST AAD6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X83226; CA58227.1; -.
DR EMBL: Z71607; CA96264.1; -.
KW SGL; S0005275; AAD14.
KW Oxidoreductase.
FT ACT_SITE 151
SQ SEQUENCE 376 AA; 41991 MW; 2265406386938313 CRC64;

```

Query Match 15.8%; Score 273; DB 1; Length 376;  
 Best Local Similarity 27.0%; Pred. No. 2.3e-14;  
 Matches 86; Conservative 51; Mismatches 142; Indels 40; Gaps 8;

```

OY      22 WLGTVNFSGRVEDDALRLMDHARDGNCIDTDMWYGRMLYKHTHELINQWRLAOGGR 81
Db      43 WSG---FMGSMNKEQAFELIDAFYFAGAGNCIDTANYSY-----QNESEIWIQEWMA-SRKL 94
OY      82 REDTVLATKVGSGMSFRVNDGSLA-----RHIIASCEGSLRLGVHDIDVYOMHHHD 134
Db      95 RDQIVATKFTGDKRYKEVGGKSNVCGNHRKSLHVSVDSTKRLQDTWIDILVYHMDD 154
OY      135 RSPAPWDEWQAMDSLVAASKVSYGSSNFAGMHIAAOENARRHSLGWSHOCLYNLA 194

```

```

Db 155 YMSSTEEWMDSHLIVQOGKVLTVSDTPAWVVSAAANYATSHGKTPPSVYQGMWNLN 214
Oy 195 RHAEELEVPAAQAYGLGYFAMSPHLGGLSLGALKERLAAGTAVKASQGRAQVLLPRLAPAI 254
Db 215 RQFEDDIIPMARHFGMALAPMDVMGGRFQ---SKKAMEERKKNEGS-----LRFV 263
Oy 255 -----EAYEFCRNMG-EDPAEVLAVLSRPGIAGAVIGPRPEQLDALK 300
Db 264 GGEQTELEVKISALTKIAEEHGTESVTAIAIAYRSKAKNVFPLIGRKIEHKONIE 323
Oy 301 ASAMTLDQALSELDEIFP 319
Db 324 ALSIKLPPEQILEYESTVP 342

RESULT 12
AAD_PHACH STANDARD: PRT: 385 AA.
AC 001752:
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 35, Last sequence update)
DE ARYL-ALCOHOL DEHYDROGENASE [NMDP+] (EC 1.1.1.91) (AAD).
OS Phanerochaete chrysosporium.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Aphylliphorales;
OC Corticiaceae; Phanerochaete.
OX NCBI_TaxID=5306;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BKM-F-1767:
RX MEDLINE=95050593; PubMed=7961751;
RA Reiser J., Muhlem A., Hardegger M., Frank G., Fiechter A.;
RT "Aryl-alcohol dehydrogenase from the white-rot fungus Phanerochaete
chrysosporium. Gene cloning, sequence analysis, expression, and
RT purification of the recombinant enzyme.";
RL J. Biol. Chem. 269:28152-28159(1994).
CC -I- CATALYTIC ACTIVITY: AN AROMATIC ALCOHOL + NADP(+) = AN AROMATIC
CC ALDEHYDE + NADPH.
CC -I- PPM: THE N-TERMINUS IS BLOCKED.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L08964; AAA61931.1; -
RW Aromatic hydrocarbons catabolism; Oxidoreductase; NADP.
FT ACT_SITE 153 153 HYDROGEN-BOND DONOR (POTENTIAL).
SQ SEQUENCE 385 AA; 43563 MW; 7D5DEB0069B9A20 CRC64;

Query Match 15.5%; Score 268; DB 1; Length 385;
Best Local Similarity 24.0%; Pred. NO. 5.9e-14;
Matches 82; Conservative 73; Mismatches 147; Indels 40; Gaps 8;

Oy 10 RLGR-----SALLTSRLMIGTVNFS-----GRVEDDALRLMDHARDRGINCIDT 54
Db 13 KLGRHQLAPGCCGLHVSITQLGAMSTIGDKMHPYGMGTMDKESFLDLDAFYVAGNFTDT 72
Oy 55 ADMYGRLYKGTTELVRWLVAOAGGRREDTVLATKVG-----GEMSERVNDGSL 105
Db 73 ANVY-----QDESEEFIEEMW-EARGNRDQWVAATKYSLVYKRGASFBEIPKITYGNS 127
Oy 106 ARHITASEGSLRLGLVDHYDVOHMHIDRSAPMDEVWQAMDSLVASKVSTVGSSNPA 165
Db 128 LKSMHISVHDSLRKLRYSIDIFYVHFWDYCTIEVVMKGNHLNVAOGKVLTVGYSDTPA 187
Oy 166 WHIAAOGENAARRHSLGVSHQCLTNLAVRHAEELEVPAAQAYGLGYFAMSPHLGGLSLG 225

```

```

Db 168 WYVSKANNYARNAGKTPEVITYEGEMNITWRDMEKDIIPICCHEGALAPWNLCAKGIKRT 247
Oy 226 ALEKLAAGTAVASAGGRAQVLL-----PSLRPIAEAYEKFCRNMG-EDPAEVLAVWL 277
Db 248 DAEF---ERRLKSAGGGRILLOFDGMLNRETERKSKALEKAEELGAKSTISVAIATLM 304
Oy 278 SRPGIAGAVIGPRPEQLDALKASAMTLDQALSELDEIFP 319
Db 305 QKFPYVFPVIGRKVEHLYANLEALDISLSPQOMFLNDTVP 346

RESULT 13
AAD4_YEAST STANDARD: PRT: 329 AA.
AC 007747:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE PUTATIVE ARYL-ALCOHOL DEHYDROGENASE AAD4 (EC 1.1.1.-).
GN AAD4 OR YDL243C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Alt-Moerbe J., Schneider C., Moro M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z74291; CAA98823.1; -
RW Oxidoreductase.
FT ACT_SITE 105 105 HYDROGEN-BOND DONOR (POTENTIAL).
SQ SEQUENCE 329 AA; 36977 MW; 62BDBD0C889610E0 CRC64;

Query Match 15.2%; Score 263.5; DB 1; Length 329;
Best Local Similarity 25.6%; Pred. NO. 1.1e-13;
Matches 77; Conservative 60; Mismatches 147; Indels 17; Gaps 6;

Oy 30 GRVEDDALRLMDHARDRGINCIDTADMYGRLYKGTTELVRWLVAOAGGRREDTVLAT 89
Db 2 GSMNEQAFELLDAYFEGAGNCIDTANSY---QNESEIWIWEMW-KSRKLRDQIVAT 56
Oy 90 KYVGEMSERVNDGSLSAH-----IIASCGSLRLGLVDHYDVOHMHIDRSAPMDEV 142
Db 57 KFTGGKYEVGGGSKSANVCNKHKSHLVSRDLSRLQTDIDILYVHMWYMSISIEV 116
Oy 143 WQAMDSLVASGVSYVSSNFAGMWHIAAOGENAARRHSLGAMVSHOCLYVLAIRHAEELEVL 202
Db 117 MDSLHIIIVQOGKVLTVGYSDTPAWVVSAAANYATSHGKTPPSVYQGMWNLNDRDERDI 176
Oy 203 PAAQAYGLGYFAMSPHLGGLSGAL--EKLAAGTAVKASQGRAQVLLPSLRPAIAYEK 259
Db 177 PMAHFHGMALAPMDVMGGRFQSKRAMERKKNKGSLRTVSGTSKQTPDEVKIS-EALAK 235
Oy 260 FCNRNG-EDPAEVLAVLSRPGIAGAVIGPRPEQLDALKASAMTLDQALSELDEIFP 318
Db 236 VAEHGTESVTAIAIAYRSKAKNVFPLVGKRIEHLKONIEALSIKLPPEQILEYESTI 295
Oy 319 P 319
Db 296 P 296

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:23:25 ; Search time 19.19 Seconds  
(without alignments)  
1321.841 Million cell updates/sec

Title: US-09-463-705A-2

Perfect score: 1728

Sequence: 1 MTTDAATVRCGRSALLTSR.....LDEIRPAVASGGAPEANLQ 333

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :

1: PIR68:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	804.5	46.6	329	2 E82644	sugar-phosphate de
2	791.5	45.8	362	2 D82644	sugar-phosphate de
3	763	44.2	316	2 F69978	sugar-phosphate de
4	495	28.6	354	2 T44934	moca protein (impo
5	472.5	27.3	333	2 H83427	probable oxidoredu
6	418	24.2	324	2 E85538	probable MAD(P)H-d
7	418	24.2	368	2 C64771	probable oxidoredu
8	414.5	24.0	352	2 C82294	oxidoreductase (as
9	408	23.6	316	2 F83170	probable oxidoredu
10	399	23.1	324	2 D84315	aryl-alcohol dehyd
11	393	22.7	344	2 T41659	probable potassium
12	384	22.2	315	2 A75289	probable potassium
13	380.5	22.0	328	2 T52133	potassium channel
14	373	21.6	345	2 B83093	probable oxidoredu
15	371.5	21.5	342	2 S61978	hypothetical prote
16	363	21.0	419	2 I55463	K+ channel beta-su
17	360.5	20.9	330	2 T07394	potassium channel
18	359.5	20.8	408	2 I59393	potassium channel
19	356	20.6	401	2 S66503	potassium channel
20	355	20.5	337	2 E75257	aldo/keto reductas
21	340	19.7	319	2 H72394	K+ channel, beta s
22	339.5	19.6	336	2 T44988	oxidoreductase (im
23	338.5	19.6	351	2 T50285	probable oxidoredu
24	338	19.6	336	2 C84256	probable oxidoredu
25	337.5	19.5	387	2 F82815	voltage-gated pota
26	334	19.3	346	2 C65066	hypothetical prote
27	333	19.3	453	2 C86176	hypothetical prote
28	328	19.0	367	2 A53131	alpha-dendrotoxin-
29	328	19.0	367	2 S45312	Rckbeta2 protein -

30	326	18.9	346	2 F85935	hypothetical prote
31	325	18.8	367	2 S66502	potassium channel
32	324	18.8	329	2 C83506	probable oxidoredu
33	317.5	18.4	340	2 B96632	hypothetical prote
34	312	18.1	331	2 H83328	probable oxidoredu
35	307	17.8	307	2 S16390	auxin-induced prot
36	306	17.7	404	2 S68409	potassium channel
37	303.5	17.6	310	2 B69755	ion channel homolo
38	302.5	17.5	346	2 G65086	hypothetical prote
39	300.5	17.4	345	2 C96632	hypothetical prote
40	300	17.4	335	2 G72546	probable K+ channe
41	293.5	17.0	331	2 D69826	aldo/keto reductas
42	292.5	16.9	346	2 F85959	probable reductase
43	291.5	16.9	310	2 T35825	probable oxidoredu
44	289.5	16.8	333	2 H72307	oxidoreductase, al
45	288	16.7	338	2 T12582	auxin-induced prot

#### ALIGNMENTS

##### RESULT 1

E82644

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 02-Sep-2000

C:Accession: E82644

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MIMD:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: E82644

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-329 <SIM>

A:Cross-references: GB:AE003996; GB:AE003849; NID:9106790; PIDN:AAE84533.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

B:Riones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurume, E.E.; La

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

M.; Tsunako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: Xf1724

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 46.6%; Score 804.5; DB 2; Length 329;

Best Local Similarity 48.8%; Pred. No. 8.2e-56;

Matches 161; Conservative 52; Mismatches 106; Indels 11; Gaps 3;

QY	8	HVRGRSALLTSRLMGLTVNFSGRVEDDRLMDHARDRCINCLDTRADMGWRLK---	64
DB	3	YTHGRGLKVRSLATGMNNGELTFSKIMDTALDAINFEDPADVAGGQPTDMP	62
QY	65	---GHEELVGRWLAAGGREDFTVATKVGEMSERVNDGSLRHITACEGSLRLG	121
DB	63	KGFETSEYIGNWLAQDKSRDKTYLAKYQPMETGENDYLLSYHRRACEASLRLK	122
QY	122	VDHIDVOMHHIDISAPWDEWQAMDSLVASGKSYVSSNFAGMHTAAQENARRSL	181
DB	123	TDHIDLQMHHDIDATWETWQAFQDLIRGKLTYYGSSNFAQMDLALACQTAASRL	182



Query Match	27.3%;	Score 472.5;	DB 2;	Length 323;
Best Local Similarity	36.2%;	Pred. No. 9.5e-30;		
Matches 118: Conservative	57;	Mismatches 116;	Indels 35;	Gaps

```

      :| : ||: | |||||:||||| :||
Db 237 KNLKESDENDAOI-----AERLTGVSEELCATRAQVALAWLSKPGIAPIICTSR 288

```

Db 289 EEOLELLNAVDITLKPEQIAELE 312

RESULT 7  
C64771

Probable oxidoreductase (EC 1.-.-.) yajO - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 18-Feb-2000

C:Accession: C64771

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

-A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: C64771

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-348 <BLAT>

A:Cross-references: GB:AE00048; GB:U00096; NID:91786614; PIDN:AC73522.1; PID:91786621;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yajO

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

C:Keywords: oxidoreductase

F:151/Active site: His #status predicted

Query Match 24.2%; Score 418; DB 2; Length 348;

Best Local Similarity 34.6%; Pred. No. 26-25;

Matches 112; Conservative 52; Mismatches 124; Indels 36; Gaps 7;

QY 11 IGRSALLSRMLGTVNF-----SGVEDDRLRLMDHARGINCLDTADMGRL 62

Db 30 IKTTLERVRRLCGCTGFEPPRGHNAWTLPESSRPRIKRALEGINFEFQINFSY--- 86

QY 63 YKGFTEELVGRMLAOGGGRREDTVLATKVGESERVND--SGLSRHIIASCEGSLRL 120

Db 87 -GGSSEIYGRAL-RDFARREDVYATKV----FHRVGLPEGLSRAQLRSLDLSRL 140

QY 121 GVDHIDVYOMHHIDRSAPWDEWQAMDSLVAAGSVYSSNFGMHIAAENARRHS 180

Db 141 GMDYVILIGIRHDYNTPIEELTLEALNDVYKAGKARYTASSMHAQFOALELQKHGM 200

QY 181 LGMVSHQCILYNLAHAELEVPAAQAYGLGVFVWSPILGGLSG-----ALEKLA 231

Db 201 AOFVSAQGRVOLLPSLRPAIEAYEKFCRNIGEPFAEVLAWVSRPGIAGAVIGPRT 291

QY 232 AGTAVSAQGRVOLLPSLRPAIEAYEKFCRNIGEPFAEVLAWVSRPGIAGAVIGPRT 291

Db 261 KNLKESDENDQI-----AERLTGVEELGATRAQVALAMLISKGIAPITGTSR 312

QY 292 PRODLSALKASAMTIDEQALSEL 315

Db 313 EEOLELLNAVDITLKPEQIAELE 336

RESULT 8  
C82294  
oxidoreductase Tas, aldo/keto reductase family VC0667 [imported] - Vibrio cholerae (stra

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: C82294

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, F

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: C82294

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-352 <HEI>

A:Cross-references: GB:AE004153; GB:AE003852; NID:9965103; PIDN:AAF93832.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0667

A:Map position: 1

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 24.0%; Score 414.5; DB 2; Length 352;

Best Local Similarity 34.4%; Pred. No. 3.8e-25;

Matches 116; Conservative 56; Mismatches 140; Indels 25; Gaps 7;

QY 5 AATHVRLGSRALLTSRLMGTNFSRGVEDDRLRLMDHARGINCLDTADMG---GRR 61

Db 8 AAOYKTLPHSSLEISKICGTMTFGONSQAFOOLDVLEGVNFDIAEYVPVPTA 67

QY 62 LYKGFTEELVGRMLAOGGGRREDTVLATKVGES--ERNDS--GLSARHIIASCEGSLRL 119

Db 68 QYQKKEEETIGNWLK-SGRKEIVATVAGPRNVPYIRDKMALDHRNIIHQAQVDSLR 126

QY 120 LGVDHIDVYOMHHIDRSA-----PMD-----EVMQAMDSLVAAGSVYSSN 162

Db 127 LQTDYIDLQVLMHPQROTNTFGQLNPPYDQKEVTLTLEALNDLVBMGKRYIGVSN 186

QY 163 FAGMHIAAENARRHSLSGMVSHQCILYNLAHAELEVPAAQAYGLGVFVWSPILGGL 222

Db 187 ETPWGVSYRLAEKHELPRIVISIQNPYMLNNSFEVGLAETSHLEGVLLVSPILAFRA 246

QY 223 LSGALEKLA--AGTAVSAQGRVOLLPSLRPAIEAYEKFCRNIGEPFAEVLAWVSRP 280

Db 247 LSGKYLNGARPAARCTLQGRSRYTTEGILATEYVLAQOFGIDPQAMALAFVNPQR 306

QY 281 GIAGAVIGPRTPEOLDLSALKASAMTIDEQALSEL 317

Db 307 FVANSNIGATTEQLKNSLDSISLAELOKIOEI 343

RESULT 9  
F83170  
Probable oxidoreductase PA3795 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: F83170

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; L

., Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: F83170

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <STO>

A:Cross-references: GB:AE004798; GB:AE004091; NID:99949963; PIDN:AA07182.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3795

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 23.6%; Score 408; DB 2; Length 316;

Best Local Similarity 33.9%; Pred. No. 1.1e-24;

Matches 112; Conservative 52; Mismatches 132; Indels 34; Gaps 11;

QY 1 MTTDAATHRLGSRALLTSRLMGTNFSRGVEDDRLRLMDHARGINCLDTADMG 60

Db 1 MTR-----KLGSGLEITLVGVGNVFGMTADESTFLLDLALDAGLNCIDTAIVYS- 54

QY 61 RLYKGF-----TEELVGRMLAOGGGRREDTVLATKVGESERVNDSGLSARHIIASCEGS 116

Db 55 RWPVGHGGESEFTLIGKMLKR-TGKRDRMYIASKVGMQDN--GHRGLSAAYTEQLERS 111

QY 117 LRLVGDHIDVYOMHHIDRSAPWDEWQAMDSLVAAGSVYSSNFGMHIAAENNA 176

Db 117 LRLVGDHIDVYOMHHIDRSAPWDEWQAMDSLVAAGSVYSSNFGMHIAAENNA 176



Db 112 LRRLOTVDLYQSHDDPHFDLEETLSTYGEELIKKKGVIRIGASNDARLLAAROYSA 171  
 QY 177 RRHSIGMWSHOL---YNLAVR-HAELEVLPAQAQVAGVAFWMSPLHGGSLSGALEK--- 229  
 Db 172 R---LNLPSYSGLOPEYXNLYDRADETNLEPVELGIGVISYSLAGFSGKRNQAD 228  
 QY 230 ---LAAQTAIVKS-AAGRAQVLLPSLRPAIEAEKFCRNLSGEPAPVGLAWVSRPGIAGA 285  
 Db 229 TAGRARGEKVKCYLNERGVAILIAALDEVAEOY-----NANPVOYALAMLIARPTTAP 281  
 QY 286 VIGPRTPEQLDSALKASAMTLDEQALSEL 315  
 Db 282 IASATSLLEQLDLDLIAATHLIKIDEQALIP 311

RESULT 10  
 D84315  
 aryl-alcohol dehydrogenase [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: D84315  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: D84315  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-324 <STO>  
 A:Cross-references: GB:AE004437; NID:910581102; PIDN:AAG19888.1; GSPDB:GN00138  
 A:Genetics:  
 C:Gene: aad  
 C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 23.1%; Score 399; DB 2; Length 324;  
 Best Local Similarity 35.7%; Pred. No. 5.6e-24;  
 Matches 114; Conservative 40; Mismatches 143; Indels 22; Gaps 8;

QY 1 MTDTAATVVRIGRSALLTSRLMIGTVNF-----SGRVE--DDDAALMDHARDRCINCLD 53  
 Db 1 MGLDA---VPLGRTGLRVTELAFTGFRGRETADGDIIVGQRARELLDAVANAGGRPID 57  
 QY 54 TADWGMRLYKGTETELVGRMLAOGGRRPDTVLATKVGESSEVNDGSLARHIIASC 113  
 Db 58 TADWGMRLYKGTETELVGRMLAOGGRRPDTVLATKVGESSEVNDGSLARHIIASC 113  
 QY 114 EGSRLRLGVHDIDYOMHNDIRSAFDEWQAMDSLVASGVKSYVSSNF--AGWHIAAA 171  
 Db 112 DNIIRLCTDIDILYIHRMDATPADALMTLDFVREGVYHYGASTFHPNARKIKAKA 171  
 QY 172 QENARHRSLSLWVSHOCLYNLAVRHAELEVLPAQAQVLLPSLRPAIEAEKFCRNLSGEPAPVGLAWVSRPGIAGA 231  
 Db 172 NEIARRSSEPTTYAOPRFNAVNEIYGNYLEMKREISGLGVAPSPSLAGGELTGKTERDA 231  
 QY 232 ---AGT-AVKSAGQRAQVLLPSLRPAIEAEKFCRNLSGEPAPVGLAWVSRPGIAGA 287  
 Db 232 EPPAGSRGATQQGFVDSYLTSPNSFDTLDAVEAVADVATRAQVSLAVLRHNDVAAAIT 291  
 QY 288 GPRTPPEQLDSALKASAMTL 306  
 Db 292 GARTPAQLRENMAATNIEL 310

RESULT 11  
 T41659  
 probable potassium channel subunit - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
 C:Accession: T41659

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
 submitted to the EMBL Data Library, May 1998  
 A:Reference number: 222008  
 A:Accession: T41659  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-344 <MOO>  
 A:Cross-references: EMBL:AL023590; PIDN:CAA19066.1; GSPDB:GN00068; SPDB:SPCC965.06  
 A:Experimental source: strain 972h-; cosmid c965  
 C:Genetics:  
 A:Gene: SPDB:SPCC965.06  
 A:Map position: 3  
 C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 22.7%; Score 393; DB 2; Length 344;  
 Best Local Similarity 31.8%; Pred. No. 1.8e-23;  
 Matches 107; Conservative 55; Mismatches 144; Indels 30; Gaps 10;

QY 5 AATHVR-----LGRSALLTSRLMIG---TVNFSGRVEDDALARLMDHARDRCINCLD 53  
 Db 4 ATTHQPKNVNPPFRFLGRSLKVAFSGLGWLTYGNEG-YDVEHTFNCLKQAWDLGINTFD 62  
 QY 54 TADWGMRLYKGTETELVGRMLAOGGRRPDTVLATKV-GEEMSERVNDGSLARHIIAS 112  
 Db 63 TAEIYS---NNSSETVMKAKIKELGMDSEVYITTKVFGGTYKLPNTGLSRHIIEG 118  
 QY 113 EGSRLRLGVHDIDYOMHNDIRSAFDEWQAMDSLVASGVKSYVSSNFAGWHIAAAQ 172  
 Db 119 LNASLKRGLRPVVDYIMHNRPPSPVMEVVAATFOLLQDCAFAFWGTSSEMAFELIHAN 178  
 QY 173 ENAARRHSLGVSHOCLYNLAVR-HAELEVLPAQAQVAGVAFWMSPLHGGSLSGAL---- 227  
 Db 179 HATKYNLIAPVADQPOYNYLTFRDHFEEKDLPLQOYIGVATVWSPKSGIITJGVNDGI 238  
 QY 228 -EKLAAQTAIVKSAGRAQVLLPSLRPAIE---AYEKFPCRNLSGEPAPVGLAWVSRPGIA 283  
 Db 239 PEGSRLSTFTSLAGQLDT--PEGKTOLDQVROISKIAEQIGATPSALAWTLNPPYVS 296  
 QY 284 GAVIGPRTPEQLDSALKASAM--TLDEQALSELDEI 317  
 Db 297 TTILGASKPEQIVENVKAVEFIDKLPTELKIDDI 332

RESULT 12  
 A75289  
 probable potassium channel, beta subunit - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: A75289  
 R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Ulterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: A75289  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-315 <WHI>  
 A:Cross-references: GB:AE002063; GB:AE000513; NID:96460121; PIDN:AAFL1861.1; PID:9646  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2317  
 A:Map position: 1  
 C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 22.2%; Score 384; DB 2; Length 315;  
 Best Local Similarity 32.4%; Pred. No. 8.2e-23;  
 Matches 103; Conservative 55; Mismatches 146; Indels 14; Gaps 8;

QY 10 RLGRSALLTSRLMIGTVNFSGRVEDDALARLMDHAR---DRCINCLDTADWGMRLYKGI 66

```

Db      5 KLGKSLKSYEVALGGMETYG--VNQDASKVGVGIYAAVDEGVNFDQADVYA----RQQ 59
QY      67 TEELVGRMLAOGGGRREDYVLAATKVGSESRNDSGLSARHIIASCESLRLGVDHID 126
Db      60 SEQLKMAVLRD--FPHHTLVLSKRVFMPMSDDVNDGSLRKHLSDSLRLGVDYD 117
QY      127 VYQMHIDRSAPWDEWQAMDSLVASGVSYSSNFAGMHIAAOENARHSLQMVSH 186
Db      118 IYFAHRYDEVPMEETIVAFDQYIRDGKALYMGTSWMPARIAQAVEPAKANGLHAPVE 177
QY      187 QCLYINLAVR-HAELEVLPAQAQVGLGVFAPMSPLHGLISGAL-EKLAAGTAVKSAQGRA- 243
Db      178 QPEYSVWRDRDVEQELIPYTEGAGIGLVWSPPLAMGLTGKYGDEGRPEGARLTKENMAG 237
QY      244 QVLLPRLPAIEAEYEFKFCNKLGEPAVGLAWYLSRPGIAGAVIGRTPEDLSALKASA 303
Db      238 SYLTEDNICKVRDLKSLADDLGCTRAQALAWLLRQKGVSVITGATKYNQIDYVKAAG 297
QY      304 WRLDEQALSELDEIF-PA 320
Db      298 VRLSDDVQRRIEDILKPA 315

```

```

RESULT 13
T52133
potassium channel beta subunit homolog [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52133
R:Spoomak, P.; Palme, K.
submitted to the EMBL Data Library, April 1998
A:Description: Potassium channel beta subunit homolog.
A:Reference number: 225974
A:Accession: T52133
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-328 <SPO>
A:Cross-References: EMBL:AF061570; PIDN:AAC15999.1
C:Genetics:
A:Gene: KV-beta1
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

```

Query Match      22.0%; Score 380.5; DB 2; Length 328;
Best Local Similarity 30.6%; Pred. No. 1.6e-22;
Matches 102; Conservative 64; Mismatches 144; Indels 23; Gaps 9;

```

```

QY      11 LGRSALLTSLRLMIGT-VNFSGRVEDDRLRMDHARDRGINCIDTDMYGWRLYKGHTEE 69
Db      6 LKRSGLKAVSLTSGAWVTGNDLVKAEKSILOCCRDHGVNFDMAEYVA---NGRAEE 61
QY      70 LVGRMLAOGGGRREDYVLAATKY--GGEMSERVNDGLSARHIIASCEGLRLGVVDHID 127
Db      62 LMGQAIRELIGWRSDIVITKIFWG--PGPNDKGLSRKHIVETGKASKSLKMDMYVDV 118
QY      128 YQMHIDRSAPWDEWQAMDSLVASGVSYSSNFAGMHIAAOENARHSLQMVSHQ 187
Db      119 LVCHRDASTPIEETVRANVYIDKGMAYFWGTSEMSAQOITFAMGAARDLDVGVIVE 178
QY      188 CLYINLAVR-HAELEVLPAQAQVGLGVFAPMSPLHGLISGAL-EKLAAGTAVKSAQGRA- 242
Db      179 PEYNMFAHRYDEVPMEETIVAFDQYIRDGKALYMGTSWMPARIAQAVEPAKANGLHAPVE 238
QY      243 A-QVLLPRLPAIEAEYEFKFCNKLGEPAVGLAWYLSRPGIAGAVIGRTPEDLSALKASA 301
Db      239 ANRSILDVLRKVSGLKPIADELGVTLADALATWCAENPWSVITGATRESQIDENMKA 298
QY      302 SAMTLEQALSE---LDEIFPAVAVSGAPEAM 331
Db      299 ----VDVILPLPIVLDKIEQVIGSKFRPEST 327

```

```

RESULT 14
B83093
probable oxidoreductase PA4434 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83093
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MOID:20437337
A:Accession: B83093
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-345 <STO>
A:Cross-References: GB:AE004857; GB:AE004091; NID:g9950654; PIDN:AA007822.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4434
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

```

Query Match      21.6%; Score 373; DB 2; Length 345;
Best Local Similarity 32.0%; Pred. No. 6.7e-22;
Matches 108; Conservative 54; Mismatches 138; Indels 38; Gaps 10;

```

```

QY      11 LGRSALLTSLRLMIGT-VNFSGRVEDDRLRMDHARDRGINCIDTDMY----GWRLYKGH 66
Db      6 LKRTDLKVASLCLGTITWQEQNSEQDAFOIARAKAGINFMDTAEYVPPAEYV-AS 64
QY      67 TEELVGRMLAOGGGRREDYVLAATKVGEMS--ERVNDGL--SARHIIASCEGLRLGV 122
Db      65 TERIIIMFRRSGD--RADWILASKIAGPCNGISHVDGKLFKFRHITVALDLSERLQT 123
QY      123 DHIDYQMH-----HIDRS-APWDEWQAMDSLVASGVSYSSNFAGM 166
Db      124 DWLDLQVLMHPERRINPFQOLGYQHGESEFTPLETLEVLDEGVRAKIRHIGLSNETPW 183
QY      167 HIAAOENARHSLGCVSHOCLYNLAVRHAELEVLPAQAQVGLGVFAPMSPLHGLISGA 226
Db      184 GMTFLRLAEERKMPRAVSIQNPYNILNRSFVGLAEIARRECGGLATSPMAFGMLSG- 242
QY      227 LEKLAGTAVKSAQGRQVLL-----PSLRPAIEAEYEFKFCNKLGEPAVGLAWYLSR 279
Db      243 --KYADG--ARPANARISLYSRFTRYTNPOAEACARVVALAREHGLEPQMALAVYTSR 298
QY      280 PGIAGAVIGRTPEDLSALKASAMTLEQALSELDEI 317
Db      299 PFVTSNIIGATSLDELETNIGSVDLRLDEVLGIDAI 336

```

```

RESULT 15
S61978
hypothetical protein YPL088w - yeast (Saccharomyces cerevisiae)
N:Alternative names: hypothetical protein LPe20w
C:Species: Saccharomyces cerevisiae
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 18-Feb-2000
C:Accession: S61978
R:Yang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A:Reference number: S61959
A:Accession: S61978
A:Molecule type: DNA
A:Residues: 1-342 <MAN>
A:Cross-References: EMBL:U43281; NID:g1151218; PIDN:AA068211.1; PID:g1151238; GSPDB:G
C:Genetics:
A:Gene: MIPS:YPL089c
A:Map position: 16L
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

[illegible]

Search completed: June 22, 2001, 09:25:12  
Job time: 107 sec

W

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: June 23, 2001, 10:26:43 ; Search time 113.8 Seconds

5512.071 Million cell updates/sec

```
Title: US-09-463-705A-1_COPY_48_1046
Perfect score: 999
Sequence: 1 ctgcaccagcctccgcgcgtcgcgccgtcgtgcat 999
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

```
Searched:      730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 14602022
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database

1:	/SID8/gcgdata/gemseq/gemseqn/NA1981.DAT *
2:	/SID8/gcgdata/gemseq/gemseqn/NA1981.DAT *
3:	/SID8/gcgdata/gemseq/gemseqn/NA1982.DAT *
4:	/SID8/gcgdata/gemseq/gemseqn/NA1983.DAT *
5:	/SID8/gcgdata/gemseq/gemseqn/NA1984.DAT *
6:	/SID8/gcgdata/gemseq/gemseqn/NA1985.DAT *
7:	/SID8/gcgdata/gemseq/gemseqn/NA1986.DAT *
8:	/SID8/gcgdata/gemseq/gemseqn/NA1987.DAT *
9:	/SID8/gcgdata/gemseq/gemseqn/NA1988.DAT *
10:	/SID8/gcgdata/gemseq/gemseqn/NA1989.DAT *
11:	/SID8/gcgdata/gemseq/gemseqn/NA1990.DAT *
12:	/SID8/gcgdata/gemseq/gemseqn/NA1991.DAT *
13:	/SID8/gcgdata/gemseq/gemseqn/NA1992.DAT *
14:	/SID8/gcgdata/gemseq/gemseqn/NA1993.DAT *
15:	/SID8/gcgdata/gemseq/gemseqn/NA1994.DAT *
16:	/SID8/gcgdata/gemseq/gemseqn/NA1995.DAT *
17:	/SID8/gcgdata/gemseq/gemseqn/NA1996.DAT *
18:	/SID8/gcgdata/gemseq/gemseqn/NA1997.DAT *
19:	/SID8/gcgdata/gemseq/gemseqn/NA1998.DAT *
20:	/SID8/gcgdata/gemseq/gemseqn/NA1999.DAT *
21:	/SID8/gcgdata/gemseq/gemseqn/NA2000.DAT *
22:	/SID8/gcgdata/gemseq/gemseqn/NA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	999	100.0	3412	20	AAK25772	S. erythraea eryth
2	999	100.0	3756	18	AAF72684	Sugar biosynthesis
3	511.6	51.2	9994	22	AAC85191	S. avermitilis 10
4	150	15.0	1512	22	AAFC641103	P. putida KT2440-e
5	128.2	12.8	1481	21	AAAC44737	zea. mays DNA frag
6	92	9.2	1032	19	AAV50486	Streptomyces clav
7	92	9.2	7193	19	AAV50431	Streptomyces clav
8	84.6	8.5	4257	19	AAV68520	The nucleotide seq
9	84.6	8.5	4257	19	AAV10362	Infected cell prot
10	83.8	8.4	114955	20	AAV53491	Human adenosine A1
11	81.8	8.2	114955	20	AAK53491	Human adenosine A1

C 12	81.2	8.1	384.9	22	AAE575705
C 13	79.8	8.0	58857	21	AAE584711
C 14	77.8	7.8	44377	18	AAE785088
C 15	77.8	7.8	44377	18	AAE804414
C 16	77.4	7.7	15872	17	AAE878155
C 17	75.8	7.6	12001	16	AAO762133
C 18	75.8	7.6	15081	21	AAE872838
C 19	75.8	7.6	43280	18	AAE804133
C 20	75.2	7.5	1886	21	AAE489268
C 21	73	7.3	3957	22	AAAO96686
C 22	72.8	7.3	3978	21	AAE557853
C 23	72.8	7.3	1224.9	21	AAE558404
C 24	72.8	7.3	18331	21	AAE558575
C 25	72.6	7.3	37836	21	AAE119922
C 26	72.2	7.2	2061	10	AAAN22408
C 27	71.8	7.2	2943	17	AAE164808
C 28	71.8	7.1	1998	17	AAO111277
C 29	71.2	7.1	5392	15	AAO6420107
C 30	71	7.1	2776	18	AAE873070
C 31	71	7.1	30001	18	AAE61016
C 32	71	7.1	110001	20	AAAO5110
C 33	70.6	7.1	11220	21	AAE87298
C 34	70.6	7.1	13613	21	AAE87318
C 35	70.6	7.1	36778	21	AAE87318
C 36	70.6	7.1	37948	21	AAE87285
C 37	70.6	7.1	38506	21	AAE75633
C 38	70.6	7.1	38506	21	AAE26001
C 39	70	7.0	104	20	AAE33162
C 40	70	7.0	4330	20	AAE233182
C 41	69	6.9	2918	21	AAE239222
C 42	69	6.9	2919	21	AAE244866
C 43	67.4	6.8	2366	12	AAO10190
C 44	67.4	6.7	671	21	AAE12227
C 45	67	6.7	833	15	AAO042030

## ALIGNMENTS

post-date

Result	1
XX	AAK25772
XX	AAK25772 standard; cDNA; 3412 BP.
XX	AAK25772;
DT	08-JUN-1999 (first entry)
XX	
DE	S.erythraea erythromycin-synthesis gene cluster eryG-eryAIII.
XX	
KW	Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;
KW	secondary metabolite; eryBII; eryCIII; eryCII; hybridisation; probe;
KW	glycosylation; macrolactone; oleandomycin; ds.
OS	Saccharopolyspora erythraea.
XX	
XX	
Key	Location/Qualifiers
FT	complement (47..1048)
FT	/*tag= a
FT	/label= ORF7
FT	/gene= "eryBII"
FT	/product= "dTDP-4-keto-L-6-deoxyhexose-2,3-reductase"
CDS	complement (1045..2310)
FT	/*tag= b
FT	/label= ORF8
FT	/gene= "eryCIII"
FT	/product= "desosaminyltransferase"
CDS	complement (2321..3406)
FT	/*tag= c
FT	/label= ORF9
FT	/gene= "eryCII"
FT	/product= "dTDP-4-keto-D-6-deoxyhexose-3,4-isomerase"
PN	

post done

XX	04-FEB-1999.
PD	
XX	
PF	21-JUL-1998; 98WO-FR01593.
XX	
PR	12-JUN-1998; 98FR-0007411.
PR	25-JUL-1997; 97FR-0009458.
XX	
PA	(HMRI ) HOECHST MARION ROUSSEL.
PI	Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC;
PI	Salah-Bey K, Fromentin C, Mendez C, Salas JA;
XX	
DR	WPI: 1999-142938/12.
DR	P-PSDB; AAW99386, AAW99387, AAW99388.
XX	
PT	New nucleic acid sequences encoding enzymes involved in macroliide
PT	biosynthesis - useful for producing hybrid secondary metabolites,
PT	particularly erythromycin analogues
XX	
PS	
XX	Claim 1; Fig 2; 221pp; French.
CC	This sequence represents the eryG-eryAIII gene cluster from the
CC	Gram-positive bacterium Saccharopolyspora erythraea which encodes
CC	enzymes involved in the production of the macroliide antibiotic
CC	erythromycin as a secondary metabolite. The erythromycin gene cluster
CC	spans approximately 53 kb and contains at least 20 open reading frames
CC	(ORF). This sequence contains the eryBII, eryCIII and eryCII genes
CC	encoding ORFs 7, 8 and 9 respectively. A DNA representing the
CC	eryAII-eryK region of the same cluster is shown in AAW25774. The genes
CC	are used to produce hybrid secondary metabolites in S.erythraea, i.e.
CC	erythromycin analogues which may have improved properties or as
CC	hybridisation probes for isolating homologous genes involved in
CC	glycosylation of macrolactones in macroliide-producing strains
CC	(specifically oleandomycin-producing strains of Streptomyces
CC	antibioticus).
XX	
QQ	Sequence 3412 BP; 424 A; 1221 C; 1259 G; 508 T; 0 other;

Query Match	100.0%	Score 999	DB 20	Length 3412
Best Local Similarity	100.0%	Pred. No. 7.6e-144		
Matches 999, Conservative	0	Mismatches 0	Indels 0	Gaps 0

[illegible]

Db	470	cgccagagcttgtagcagcgactcgtctgaggagacacatgcccacagagagctgtagcgcgagcgagctt	529
QY	481	ctccctcgcgcgcgcgagatgctggtccagaccgacgaagttcgaagagccgaacgttagagaaacctt	540
Db	530	ctccctcgcgcgcgcgagatgctggtccagaccgacgaagttcgaagagccgaacgttagagaaacctt	589
QY	541	gcccctcgcgcgagagaggtctgtcccatctgcccacacactctgtccacacagcgcgagaccggttc	600
Db	590	gcccctcgcgcgagagaggtctgtcccatctgcccacacactctgtccacacagcgcgagaccggttc	649
QY	601	gattctgctgtgcatctcgtgttagagacgtctcgaatgtgtgtctgaaagccacgacctgtcgacgaatccctc	660
Db	650	gattctgctgtgcatctcgtgttagagacgtctcgaatgtgtgtgtctgaaagccacgacctgtcgacgaatccctc	709
QY	661	gcagagagagcgatgtagtctggtccggtccgacagcccgcctgtctgtttgaagcgctcgatcatctc	720
Db	710	gcagagagagcgatgtagtctggtccggtccgacagcccgcctgtctgtttgaagcgctcgatcatctc	769
QY	721	gcccctcgacacctgtgtctgcacagacaggtgtgtccctgcgcgtctcccgccacctgtgacagacca	780
Db	770	gcccctcgacacctgtgtctgcacagacaggtgtgtccctgcgcgtctcccgccacctgtgacagacca	829
QY	781	ctctgcaccacagactcctcctcgtctgtgcacctttagtagagtcgacagccgatactgtccgcggt	840
Db	830	ctctgcaccacagactcctcctcgtctgtgcacctttagtagagtcgacagccgatactgtccgcggt	889
QY	841	gtcgagagcgattgtagtgcgcggtgtcccggtgcgtgtgtgtccatcaagtcgacagcggtctgctgtc	900
Db	890	gtcgagagcgattgtagtgcgcggtgtcccggtgcgtgtgtgtgtccatcaagtcgacagcggtctgctgtc	949
QY	901	ctcgacacggtctgcctctgaaagtctcaacgtctgcacgagacaaagaaacctcgtcgtgttagagacgacgga	960
Db	950	ctcgacacggtctgcctctgaaagtctcaacgtctgcacgagacaaagaaacctcgtcgtgttagagacgacgga	1009
QY	961	acgcccagacgcgcgaactgtgcgtcgcgcgacgtctgcgtgtgtcat	999
Db	1010	acgcccagacgcgcgaactgtgcgtcgcgcgacgtctgcgtgtgtcat	1048

```

RESULT 2
AA772684/C
ID AA772684 standard; DNA; 3756 BP.
XX
XX AA772684;
AC
XX
DT 18-SEP-1997 (first entry)
XX
DE Sugar biosynthesis gene cluster.
XX
KW Polyketide; glycosylation; eryCII; eryCIII; eryBII; L-mycarose;
KW D-desosamine; antimicrobial; antibiotic; antifungal; fungicide;
KW anticancer; cyclostatic; antineoplastic; ss.
XX
OS Saccharopolyspora erythraea.
XX
XX
XX Key Location/Qualifiers
FH 54..1136
FT CDS /*tag= a
FT /*label= eryCII
FT 1147..2412
FT /*tag= b
FT /*label= eryCIII
FT 2409..3410
FT CDS /*tag= c
FT /*label= eryBII
XX
XX MO9723630-A2.
XX
XX 03-JUL-1997.
XX
XX 23-DEC-1996; 96WO-US20238.
XX

```

PR 21-DEC-1995; 95US-0576626.  
 XX (ABBO ) ABBOTT LAB.  
 XX Donadio S, Katz L, Staver MJ, Summers RG;  
 PI WPI: 1997-351066/32.  
 XX F-PSDB; AAM19734-36.  
 DR  
 XX  
 PT New genes involved in sugar biosynthesis and attachment - used to  
 PT generate polypeptide antimicrobials etc. with altered pattern of  
 PT glycosylation  
 PS  
 PS Claim 1, Fig 4A, 85pp; English.  
 XX  
 CC 2 Polynucleotides (AAT72684 and AAT72685) represent 2 distinct clusters  
 CC of genes encoding enzymes (AAM19734-36 and AAM19737-42) involved in  
 CC sugar biosynthesis and attachment in Saccharopolyspora erythraea.  
 CC The eryB genes are involved in the biosynthesis of L-mycarose and  
 CC the eryC genes involved in the biosynthesis of D-desosamine.  
 CC Novel glycosylation-modified polypeptides are produced by  
 CC selectively altering, inactivating or augmenting these eryB and/or  
 CC eryC genes and introducing them into polypeptide-producing  
 CC microorganisms.  
 CC  
 SQ Sequence 3756 BP; 562 A; 1373 C; 1351 G; 470 T; 0 other;

Query Match 100.0%; Score 999; DB 18; Length 3756;  
 Best Local Similarity 100.0%; Pred. No. 7,5e-144;  
 Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgaacacaggtctccgagccgagccgagagccacccgaggaagatctcgtccag 60  
 DB 3407 CTGCAACACAGGCTTCCGGGCGCCGCGCGAGCGCCACCGCGGGAAGATCTGTCAG 3348  
 QY 61 ttccgacacagcctgtgtctccagggctatccgagagcctttagagcgagatccgagctg 120  
 DB 3347 TTGCGACACGCGCTCTGCTCGCCAGGGTCATCGCGAGCGCTTACGCGGAGTCGAGCTG 3288  
 QY 121 ctccgaggtctccgagccagatgagccgagccgagatccgagccgaggaacacacatgc 180  
 DB 3287 CTGCGGGGCTTCCGGGCGCCAGTACGCGCGGATGCCGCGGAGACACGCCATGC 3228  
 QY 181 gagccacacctcgccgggtctctccgagaggttcgagcagaacttcgttagcctcgat 240  
 DB 3227 GAGCCCACTCGCGCGGCTCTCGCGAGGTTGCGGCAAACTCTGTAAGCGCTCGAT 3168  
 QY 241 ccgagcgagccagaggaagcgaacacacacttcgacagccctcgccgagacttcagcgagt 300  
 DB 3167 CGCGGGGCGAGGAGGAGACACGACCTGCGACGCGCTGCGGACTTCAACCGCGGT 3108  
 QY 301 gcccgagcgagcagctctccagcgctccgctgagcagagccgctgagcgagcagcagc 360  
 DB 3107 GCCCGGGCGAGCTTCTCCAGCGCTCTCGCTGAGCAGCGCCGCTGCAACGCGCAACCA 3048  
 QY 361 gaagacgagcagccgctagagcctgagcgagccgagcagcagcagcagcagcagcagc 420  
 DB 3047 GAAGACGCGAGGCGCTAGGCGCTGCGGGGCGGACACACTCAGCTCGGCGCTGCGG 2888  
 QY 421 ccgagaggttctagagcagctgtgtggaagacatgcccaaggagatgagcgagcgagcgt 480  
 DB 2987 CGCGAGGTTGTACAGCACTGTGTGAGAGACATGCCAGGAGATGCGCGCGCGCGCTT 2928  
 QY 481 ctcccgagcgagcgagcagctgacagccgagagctcgagcagcagcagcagcagcagc 540  
 DB 2927 CTCTCTCGCGGCGCGCATGTGCAAGCCCGAGAGTTGAGAGCGACGAGTGAAGACCTT 2868  
 QY 541 gccgctgagcagagagctgttccatgagcctgacacacactcgtcccaagcgcgagcagcgtc 600  
 DB 2867 GCGGCTGGGAGAGAGGCTGTGCATGTGCGACACACTCTGTCCACAGCGCGGAGCGGTC 2808  
 QY 601 gatgtgtgcatctgtgtagatgtgattgtgtcagacgcccagcctgagcagcagcagcagc 660

DB 2807 GATGTGTGATCTGTGAGACGTCATGTGTGAGCGCCAGCGCTGCGACGATCTCCTC 2748  
 QY 661 gcagagcgagatgatlgtccgagccagacagccgagctgtgtgacgagcgtcgtcattc 720  
 DB 2747 GCAGAGGCGATGATGTGCGCGCGCGACAGCCGCTGTGTGAGCGGCTGCTCATCTC 2688  
 QY 721 gccgagcagctgtgtccagcagcagctgtctcgtcgagcgagccgagccagccca 780  
 DB 2687 GCGCGGACCTTGTGTGCGACGAGCGGTCTGCGCGCGCTGCGCGCGCGCGCGCA 2628  
 QY 781 cctgcccacagcagcctctcgtgtgtgagcagcagcagcagcagcagcagcagcagc 840  
 DB 2627 CTTGCCACACAGCTCTCTGCTGTGCGCTGTGTGAGCGCGCGACCGTACATGTGCGG 2568  
 QY 841 gtccagagcagctgtatccgagcagcagcagcagcagcagcagcagcagcagcagcagc 900  
 DB 2567 GTGAGGCACTTATGATCCCGCGGTCCGCGCGGTGTGATCAGCGCGAGCGCTGTGTC 2508  
 QY 901 ctccagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 960  
 DB 2507 CTGACGCGTGTGCTGAAATTCACCGTGTGCGGACCGACCAAGCGCTGCTGTGAG 2448  
 QY 961 acgcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 999  
 DB 2447 ACGCCGAGCGCGACGATGCTGCGGCGGTGCTGAT 2409

RESULT 3  
 AAC85191  
 ID AAC85191 standard; DNA; 9994 BP.  
 XX AAC85191;  
 AC 14-MAY-2001 (first entry)  
 DT  
 XX  
 DE S. avermitilis 10 kb genomic DNA.  
 XX  
 KW Glycosylation; avermectin; aglycone; biocatalysis; antibiotic;  
 KM oleandrose; ds.  
 OS  
 XX Streptomyces avermitilis.  
 FH  
 FT Key  
 FT CDS  
 FT Location/Qualifiers  
 FT 508..1332  
 FT /tag= a  
 FT /product= "Macrolide B-keto reductase"  
 FT /note= "ORF1 protein AAB61973"  
 FT 1390..2628  
 FT /tag= b  
 FT /product= "Glycosyl transferase"  
 FT /gene= "AvrB"  
 FT /note= "ORF2 protein AAB61974"  
 FT complement (3598..4497)  
 FT /tag= c  
 FT /product= "TDP-glucose synthase"  
 FT /gene= "AvrD"  
 FT /note= "ORF3a protein AAB61976"  
 FT complement (2534..3613)  
 FT /tag= d  
 FT /product= "TDP-glucose 4, 6 dehydratase"  
 FT /gene= "AvrC"  
 FT /note= "ORF3b protein AAB61975"  
 FT 4624..5655  
 FT /tag= e  
 FT /product= "Glycosyl reductase"  
 FT /gene= "AvrE"  
 FT /note= "ORF4 protein AAB61977"  
 FT complement (5709..6389)  
 FT /tag= f  
 FT /product= "Glycosyl 3, 5 epimerase"  
 FT /gene= "AvrF"  
 FT /note= "ORF5 protein AAB61978"





XX 27-JUL-1999; 99DE-1035088.  
 PF 27-JUL-1999; 99DE-1035088.  
 PR  
 XX  
 XX (TIGR-) TIGR INST GENOMIC RES.  
 PA (QUIA-) QUIAGEN GMBH.  
 PA (GEBR.) GES BIOTECHNOLOGISCHE FORTSCHRITTUNG MBH.  
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.  
 PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.  
 XX  
 DR WPI: 2001-192469/20.

XX New DNA sequences specific for *Pseudomonas putida* KT2440, useful as  
 PT related genetic engineering host, allow detection in presence of other  
 PT related bacteria -  
 XX  
 PS  
 XX Claim 1a; Page 155; 158pp; German.

CC This invention describes novel DNA sequences (1) for specific detection  
 CC of *Pseudomonas putida* KT2440. The invention also describes (1)  
 CC recombinant expression vector containing (1); (2) prokaryotic or  
 CC eukaryotic cells transformed or transfected with (1) or the vector of  
 CC (1); (3) production of expression products by culturing cells of (2);  
 CC (4) expression products, or their fragments, of (1) and synthetic  
 CC proteins or peptides with the same sequences (A); (5) poly- or  
 CC monoclonal antibodies (Ab) that react specifically with (A); (6)  
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic  
 CC plants that contain transformed or transfected cells of (2); (8)  
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips  
 CC carrying one or more (1), (1), and their fragments, are used as probes  
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by  
 CC polymerase chain reaction, and for production of transgenic plants. (1),  
 CC or antibodies that recognize their expression products, are used for  
 CC detecting the presence of KT2440, particularly in presence of other,  
 CC even closely related, bacteria. KT2440 is one of the bacteria classified  
 CC as safe, by the National Institutes of Health, for genetic engineering  
 CC work, e.g. as microbial production strains, for biological remediation  
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant  
 CC homology with sequences in other bacteria (specifically the closely  
 CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it  
 CC has greater catabolic activity and better survival in, and adaptation to,  
 CC the rhizosphere and soil.  
 CC  
 XX  
 SQ Sequence 1512 BP; 320 A; 452 C; 482 G; 257 T; 1 other;

Query Match 15.0%; Score 150; DB 22; Length 1512;  
 Best Local Similarity 50.1%; Pred No. 3.2e-15;  
 Matches 400; Conservative 0; Mismatches 396; Indels 3; Gaps 1;

QY 112 gtcgagctgctcgaggggttcgagcgagtgacgagcgccgagcgagcgagcgag 171  
 DB 900 gtccagagtggtggcgcttcgagcgagtgagcgagcgagcgagcgagcgag 841  
 QY 172 caaccatgtagagccaccactcgagcgaggtcttcgagcgaggttcgagcgagcgag 231  
 DB 840 caccacagcgccagcgaacttgagcgagcgagcgagcgagcgagcgagcgag 781  
 QY 232 ggcctcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 291  
 DB 780 caccatgtagagccaccactcgagcgaggtcttcgagcgaggttcgagcgagcgag 721  
 QY 292 caacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 348  
 DB 720 atcgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 661  
 QY 349 cagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 408  
 DB 660 cggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 601  
 QY 409 ggcgttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 468

DB 600 ttcttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 541  
 QY 469 ggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 528  
 DB 540 gttgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 481  
 QY 529 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 588  
 DB 480 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 421  
 QY 589 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 648  
 DB 420 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 361  
 QY 649 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 708  
 DB 360 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 301  
 QY 709 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 768  
 DB 300 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 241  
 QY 769 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 828  
 DB 240 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 181  
 QY 829 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 888  
 DB 180 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 121  
 QY 889 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 907  
 DB 120 gtagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 102

RESULT 5  
 AAC44737/c  
 ID AAC44737 standard; DNA; 1481 BP.  
 XX  
 AC AAC44737;  
 XX  
 XX 18-OCT-2000 (first entry)  
 DT  
 XX  
 DE Zea mays DNA fragment SEQ ID NO: 43945.  
 XX  
 XX Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway; metabolic;  
 KW pathway; promoter; termination sequence; corn; ss.  
 OS  
 XX Zea mays subsp. mays.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 XX 25-FEB-1999; 9905-0121825.  
 PR 05-MAR-1999; 9905-0123180.  
 PR 09-MAR-1999; 9905-0123548.  
 PR 23-MAR-1999; 9905-0125788.  
 PR 25-MAR-1999; 9905-0126264.  
 PR 29-MAR-1999; 9905-0126785.  
 PR 01-APR-1999; 9905-0127462.  
 PR 08-APR-1999; 9905-0128234.  
 PR 16-APR-1999; 9905-0128714.  
 PR 19-APR-1999; 9905-0129845.  
 PR 21-APR-1999; 9905-0130077.  
 PR 23-APR-1999; 9905-0130449.  
 PR 28-APR-1999; 9905-0130510.  
 PR 30-APR-1999; 9905-0131449.  
 PR 9905-0132048.

PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132863.  
PR 14-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134221.  
PR 18-MAY-1999; 990S-0134370.  
PR 19-MAY-1999; 990S-0134768.  
PR 20-MAY-1999; 990S-0134941.  
PR 21-MAY-1999; 990S-0135124.  
PR 24-MAY-1999; 990S-0135353.  
PR 25-MAY-1999; 990S-0135629.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139452.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.  
PR 18-JUN-1999; 990S-0139753.  
PR 21-JUN-1999; 990S-0139817.  
PR 22-JUN-1999; 990S-0139899.  
PR 23-JUN-1999; 990S-0140354.  
PR 23-JUN-1999; 990S-0140354.  
PR 24-JUN-1999; 990S-0140695.  
PR 28-JUN-1999; 990S-0140823.  
PR 29-JUN-1999; 990S-0140991.  
PR 30-JUN-1999; 990S-0141287.  
PR 01-JUL-1999; 990S-0141842.  
PR 01-JUL-1999; 990S-0142154.  
PR 02-JUL-1999; 990S-0142055.  
PR 06-JUL-1999; 990S-0142290.  
PR 08-JUL-1999; 990S-0142803.  
PR 09-JUL-1999; 990S-0142920.  
PR 12-JUL-1999; 990S-0142977.  
PR 13-JUL-1999; 990S-0143542.  
PR 14-JUL-1999; 990S-0143624.  
PR 15-JUL-1999; 990S-0144005.  
PR 16-JUL-1999; 990S-0144085.  
PR 16-JUL-1999; 990S-0144086.  
PR 19-JUL-1999; 990S-0144325.  
PR 19-JUL-1999; 990S-0144331.  
PR 19-JUL-1999; 990S-0144332.  
PR 19-JUL-1999; 990S-0144333.  
PR 19-JUL-1999; 990S-0144334.  
PR 19-JUL-1999; 990S-0144335.  
PR 20-JUL-1999; 990S-0144352.  
PR 20-JUL-1999; 990S-0144632.  
PR 20-JUL-1999; 990S-0144884.  
PR 21-JUL-1999; 990S-0144814.  
PR 21-JUL-1999; 990S-0145086.  
  
PR 21-JUL-1999; 990S-0145088.  
PR 22-JUL-1999; 990S-0145085.  
PR 22-JUL-1999; 990S-0145087.  
PR 22-JUL-1999; 990S-0145089.  
PR 22-JUL-1999; 990S-0145192.  
PR 23-JUL-1999; 990S-0145145.  
PR 23-JUL-1999; 990S-0145218.  
PR 23-JUL-1999; 990S-0145224.  
PR 26-JUL-1999; 990S-0145276.  
PR 27-JUL-1999; 990S-0145913.  
PR 27-JUL-1999; 990S-0145918.  
PR 27-JUL-1999; 990S-0145919.  
PR 28-JUL-1999; 990S-0145951.  
PR 02-AUG-1999; 990S-0146386.  
PR 02-AUG-1999; 990S-0146388.  
PR 02-AUG-1999; 990S-0146389.  
PR 03-AUG-1999; 990S-0147038.  
PR 04-AUG-1999; 990S-0147204.  
PR 04-AUG-1999; 990S-0147302.  
PR 05-AUG-1999; 990S-0147192.  
PR 05-AUG-1999; 990S-0147260.  
PR 06-AUG-1999; 990S-0147303.  
PR 06-AUG-1999; 990S-0147416.  
PR 09-AUG-1999; 990S-0147493.  
PR 09-AUG-1999; 990S-0147935.  
PR 10-AUG-1999; 990S-0148319.  
PR 11-AUG-1999; 990S-0148319.  
PR 12-AUG-1999; 990S-0148341.  
PR 13-AUG-1999; 990S-0148565.  
PR 13-AUG-1999; 990S-0148684.  
PR 16-AUG-1999; 990S-0149368.  
PR 17-AUG-1999; 990S-0149175.  
PR 18-AUG-1999; 990S-0149426.  
PR 20-AUG-1999; 990S-0149722.  
PR 20-AUG-1999; 990S-0149723.  
PR 22-AUG-1999; 990S-0149929.  
PR 23-AUG-1999; 990S-0149902.  
PR 23-AUG-1999; 990S-0149930.  
PR 25-AUG-1999; 990S-0150566.  
PR 26-AUG-1999; 990S-0150884.  
PR 27-AUG-1999; 990S-0151065.  
PR 27-AUG-1999; 990S-0151066.  
PR 27-AUG-1999; 990S-0151080.  
PR 27-AUG-1999; 990S-0151080.  
PR 30-AUG-1999; 990S-0151303.  
PR 31-AUG-1999; 990S-0151338.  
PR 01-SEP-1999; 990S-0151930.  
PR 07-SEP-1999; 990S-0152363.  
PR 10-SEP-1999; 990S-0153070.  
PR 13-SEP-1999; 990S-0153758.  
PR 15-SEP-1999; 990S-0154018.  
PR 16-SEP-1999; 990S-0154039.  
PR 20-SEP-1999; 990S-0154779.  
PR 22-SEP-1999; 990S-0155139.  
PR 23-SEP-1999; 990S-0155486.  
PR 24-SEP-1999; 990S-0155569.  
PR 28-SEP-1999; 990S-0156458.  
PR 29-SEP-1999; 990S-0156596.  
PR 04-OCT-1999; 990S-0157117.  
PR 05-OCT-1999; 990S-0157753.  
PR 06-OCT-1999; 990S-0157865.  
PR 07-OCT-1999; 990S-0158029.  
PR 08-OCT-1999; 990S-0158232.  
PR 12-OCT-1999; 990S-0158369.  
PR 13-OCT-1999; 990S-0159293.  
PR 13-OCT-1999; 990S-0159294.  
PR 13-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159295.  
PR 14-OCT-1999; 990S-0159329.  
PR 14-OCT-1999; 990S-0159331.  
PR 14-OCT-1999; 990S-0159337.  
PR 14-OCT-1999; 990S-0159637.  
PR 18-OCT-1999; 990S-0159638.  
PR 18-OCT-1999; 990S-0159584.  
PR 21-OCT-1999; 990S-0160741.

```

PR 21-OCT-1998; 990S-0160767.
PR 21-OCT-1998; 990S-0160768.
PR 21-OCT-1998; 990S-0160770.
PR 21-OCT-1998; 990S-0160814.
PR 21-OCT-1998; 990S-0160815.
PR 22-OCT-1998; 990S-0160880.
PR 22-OCT-1998; 990S-0160981.
PR 22-OCT-1998; 990S-0160989.
PR 25-OCT-1998; 990S-0161404.
PR 25-OCT-1998; 990S-0161405.
PR 25-OCT-1998; 990S-0161406.
PR 26-OCT-1998; 990S-0161359.
PR 26-OCT-1998; 990S-0161360.
PR 26-OCT-1998; 990S-0161361.
PR 28-OCT-1998; 990S-0161920.
PR 28-OCT-1998; 990S-0161921.
PR 28-OCT-1998; 990S-0161922.
PR 28-OCT-1998; 990S-0161923.
PR 29-OCT-1998; 990S-0162142.

```

Query Match 12.8%; Score 128.2; DB 21; Length 1481;  
 Best Local Similarity 55.5%; Pred. No. 6,4e-12;  
 Matches 317; Conservative 1; Mismatches 235; Indels 18; Gaps 3;

```

QY 412 gtgcgcagccagccagctgtacagcagctgtgagagaccacagccagagagtgagcg 471
DB 714 GTGCTGGAGAAAGTTTACTCCGGCTGCTCGACGATGGGTCGACAGGTCCAGAGCGG 655
QY 472 ggcgcgctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 531
DB 654 GTTGGGAGATGCGCCAGGCTCGGTGATCTGCGCGGACCACTCGAGAGTGGCCAGTA 595
QY 532 ggaagaccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 591
DB 594 GAAGGCGCCAGCCTGCTGATGACCCAGTTGATGGCGCGCACCGTCTCTCGACGGGCGT 535
QY 592 ggaacgcctcagatgtgcatcgtgtagacgtcagatgtgtagacgcgcgcgcgcgcgc 651
DB 534 GGAGGCGCTGGGGCGGCTGGCAGTACGACGCTCGACGTGATGCTCATGTCAGCGCGG 475
QY 652 cgaacccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 711
DB 474 CGAGCGCGGAGGCGCTCGACGAGCTGTCGGGAGAGGCGCCCTTGTGTTGGGCGCTG 415
QY 712 gtcacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 771
DB 414 GC---CGCCCGAAGAGCTGTGTGAGTGAACGACGTGGAAGCGCGCCACCCAGGTC 358
QY 772 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 831
DB 357 CCGGATGGGCTGGCCCATGATCTCTCGGCGGCGGCGCTT-----GGCGTACAC 310
QY 832 gtgcgcgctgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 891
DB 309 CTCGGCGTGTGAGAGAGTGTGACGCCCGCTCGGCGACGCGCTGACAGCGCTTGGC 250
QY 892 gtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 948
DB 249 CTCCTTCACGTCCACCTGTTCCCGAAGGTACCCACGCGCGTACGAGAGCTGCTGAC 190
QY 949 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 979
DB 189 GCGGAGCCCGGACCGCGCCAGGTTCTTGTAC 159

```

RESULT 6  
 AAV50486/c  
 ID AAV50486 standard; DNA; 1032 BP.  
 AC AAV50486;  
 XX 26-OCT-1998 (first entry)  
 DT Streptomyces clavuligerus DNA sequence of orfup1.  
 XX

```

XX XX Streptomyces clavuligerus; bacterial gene: clavulanic acid; 5S clavum;
KW 5R clavum; cas1; ORF; open reading frame; biosynthesis; ss.
XX XX Streptomyces clavuligerus.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..1032
FT /tag= a
FT /product= "orfup1 protein"
FT /note= "no stop codon given"
XX
XX MO9833896-A2.
XX
XX 06-AUG-1998.
XX
XX 02-FEB-1998; 98WO-EP00644.
XX
XX 04-FEB-1997; 97GB-0002218.
XX
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX (UYAL-) UNITV ALBERTA.
XX
XX Anders C, Barton B, Griffin JP, Jensen S, Mosher RH;
XX Paraskar AS;
XX WPI. 1998-437451/37.
XX P-PSDB; AAM69711.
XX
XX DNA comprising defective 5S clavulin biosynthesis gene(s) from
XX Streptomyces clavum - useful for producing clavulanic acid without
XX production of 5S clavum or clavam-2-carboxylates
XX
XX Claim 3; page 15-24; 29pp; English.
XX
XX The present sequence represents DNA comprising a gene specific for 5S
XX clavulin biosynthesis from Streptomyces clavuligerus (SC) and which is
XX not essential for 5R clavulin biosynthesis. The present invention also
XX describes: (1) a process for improving 5R clavulin production in a
XX suitable microorganism comprising manipulation of DNA as above and its
XX inclusion in the microorganism; (2) a process for improving 5R clavulin
XX production in SC comprising disrupting or otherwise making defective DNA
XX regions flanking cas 1; (3) a process for the identification of a
XX microorganism suitable for high 5R clavulin production comprising a
XX preliminary screening for microorganisms with low or no 5S clavulin
XX production; (4) a microorganism which is capable of 5R clavulin production
XX and low or no 5S clavulin production obtainable by a process as in (3);
XX (5) clavulanic acid (CA) obtainable by the fermentation of a
XX microorganism as in (4); and (6) CA which is free of any 5S clavulin; (7)
XX CA which is free of any clavam-2-carboxylate. The methods and products
XX can be used to produce organisms capable of producing increased amounts
XX of clavams suitably e.g. CA, for use as antibiotics. The methods can
XX also be used for the production of CA without the production of 5S
XX clavum or clavam-2-carboxylate.
XX
XX Sequence 1032 BP; 119 A; 394 C; 372 G; 147 T; 0 other;

```

Query Match 9.2%; Score 92; DB 19; Length 1032;  
 Best Local Similarity 47.6%; Pred. No. 2e-06;  
 Matches 438; Conservative 0; Mismatches 440; Indels 42; Gaps 4;

```

QY 27 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 86
DB 994 CCGCGACACCGGATGAGATCGCGCGCTACCTCCGAGAGGACCGGTCTCCAGCC 935
QY 87 tcatcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 146
DB 934 GATTCAGGCGGCGGCGGATGTTCCGCGAGATGGCGGCTTCCGCGGATTCG 875
QY 147 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 206
DB 874 GGACGACGTCCTCGCCCGGTGTCAGCGACGCGGCGGAGCTGTGCGAGGTCAGGC 815

```

OY	207	cgaggttcgagcaactctctcgaagccctcgaatcgccggagggc-----attgagc	257
Db	814	CCAGACCGTCCGGCACCACCGGGGACAGCGGGGTGCACCAACAGACGGTTGCGGGACAGGGCGG	755
OY	258	gcaaacagacactctgcacagcgccctctgcgcgaactcaacgcggtgtcccgggcagacttct	317
Db	754	GAGCGCTGAAACCGGGGGGCGCCCGGGGGAAGTCTCTGTCCTCCCAATCTCTCGGTGTGC	695
OY	318	ccagagcttcgcgtatgcagagcgccgcgtgtgcagcgcgcgaaccagcggaagcgcgagccgt	377
Db	694	GGATGTGTCCTCCGGTAAAGAAAACCCCGTCCAGAGGGGGGTAAAGACGATGCCATGCCCA	635
OY	378	aggcctgcgcgcgcgcgaaccctcccaagctctcgcgcgtgcgcgaacgcaggtttatagcg	437
Db	634	GCTCCGGGCAACAGGGGACACCACTCTGCTTCGTGATCCGGCCGACACAGGCTCCACTCCG	575
OY	438	actggttggaaacatctgccagggagtgtgcgcgcgcgcgcgcgttctcctctgcgcgcgcga	497
Db	574	TCTG-----CAGCGCGTACACCGGGGTGACACCGGTGCTCCCGCGGAGGAGTGCGCGG	521
OY	498	tgtgcagaccgcgcgaagttgcagagcgacgcgaagtatgagaaccttgcgcgtgcgcgcagagc	557
Db	520	AGGGGTGC-----GAGAGACCGAGCCTGCGGACCTTGCCCTCGCGCACCAAGCT	473
OY	558	tgtcattgaccttcacacactcgtcccaagcgcgcgagccgcgttcgattgttgtatctgt	617
Db	472	CGGCGCACCGCACCCACGCGTCTCTCTCATTCGGCACCGCGGGTCTCTCATGCTGTAGT	413
OY	618	agacgtcgaattgtctgcagcgccacgccttcgcgcagcagatccctctgcgcgcgcgcgatgt	677
Db	412	ACAGGTGCGATCGGTGCGGTGCGCGAGAGCAGAGGACGCTTGCAAGGCGCGGAGCGT	353
OY	678	gcgcgcgcgcgaagccgcgttgtgtgcgcgcgtctgcctcatctctgcgcgcgcgcgcgtgtcg	737
Db	352	AGGACGG-----TCGCGCAACAAGCCTTGGAAGGGCGCGTCCGACGAGC	308
OY	738	ccagcagcgttctctcgcgcgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	797
Db	307	GCACATGTCGCAATTGTTGGCATTACAGCACTGTCGCCGGCGCCGACCGCCCGT	248
OY	798	cgcgtgtgcacctgtatagccgcgcgcgcgcgcgtacatgtctgcgcgtgtctgcgcgcgcgcgc	857
Db	247	CGACAGGTCCTCTACCGCGCGCCGACCCCTTGAGACGTGCGCGGTGTCAGAGGAGTGAGCC	188
OY	858	cgcgctcccgcggtgtgtcatcagc	917
Db	187	CGCGGTGACAGCGCGCGCGGTGCGTGCCTCCCGGGCGCGGTCCGGCGCTCGTAGA	128
OY	918	agttaccagtgc	937
Db	127	AGTCGTGTCGCGCAGGCGAG	108
RESULT	7		
AAVS0431			
ID	AAVS0431	standard; DNA: 7193 BP.	
XX			
XX	AAVS0431:		
XX			
XX	26-OCT-1998	(first entry)	
DE		Streptomyces clavuligerus DNA sequence of ORFs flanking cas1.	
XX			
KW		Streptomyces clavuligerus; bacterial gene; clavulanic acid; 5S clavum	
KW	5R	clavum; cas1; ORF; open reading frame; biosynthesis; ss.	
XX			
OS		Streptomyces clavuligerus.	
XX			
XX	MO9833896-A2.		
XX			
XX	06-AUG-1998.		
XX			

PF 02-FEB-1998; 98WO-EP00644.  
PR 04-FEB-1997; 97GB-0002218.  
XX (SMIK ) SMITHKLINE BEECHAM PLC.  
PA (UYAL-) UNIV ALBERTA.  
XX  
PI Anders C, Barton B, Griffin JP, Jensen S, Mosher RH;  
PI Parakkhar AS;  
XX  
XX WPI: 1998-437451/37.  
DR  
PT DNA comprising defective 5S clavin biosynthesis gene(s) from  
PT Streptomyces clavam - useful for producing clavulanic acid without  
PT production of 5S clavam or clavam-2-carboxylates  
XX  
XX Claim 2; Page 15-24; 29pp; English.  
XX  
CC The present sequence represents DNA comprising genes specific for 5S  
CC clavam biosynthesis from Streptomyces clavuligerus (SC) and which are  
CC not essential for 5S clavam biosynthesis. The present invention also  
CC describes: (1) a process for improving 5R clavam production in a  
CC suitable microorganism comprising manipulation of DNA as above and its  
CC inclusion in the microorganism; (2) a process for improving 5R clavam  
CC production in SC comprising disrupting or otherwise making defective DNA  
CC regions flanking cas 1; (3) a process for the identification of a  
CC microorganism suitable for high 5R clavam production comprising a  
CC preliminary screening for microorganisms with low or no 5S clavam  
CC production; (4) a microorganism which is capable of 5R clavam production  
CC and low or no 5S clavam production obtainable by a process as in (3);  
CC (5) clavulanic acid (CA) obtainable by the fermentation of a  
CC microorganism as in (4), and (6) CA which is free of any 5S clavam; (7)  
CC CA which is free of any clavam-2-carboxylate. The methods and products  
CC can be used to produce organisms capable of producing increased amounts  
CC of clavams suitably e.g. CA, for use as antibiotics. The methods can  
CC also be used for the production of CA without the production of 5S  
CC clavam or clavam-2-carboxylate.  
XX  
XX Sequence 7193 BP; 1055 A; 2745 C; 2498 G; 895 T; 0 other;

[illegible]

[illegible]

Result	ID	AAV68520/C	AAV68520 standard; DNA; 4257 BP.
XX	AC	AAV68520;	
XX	DT	29-JAN-1999	(first entry)
XX	DE	The nucleotide sequence of the Herpes simplex virus ICP4.	
XX	KW	Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;	
XX	OS	stimulation; inhibition; HSV infection; ss.	
XX	FT	Herpes simplex virus.	
XX	FT	Key	Location/Qualifiers
XX	FT	CDS	361..4257
XX	FT		/*tag= a
XX	PN	WO9846637-A2.	/product= "HSV ICP4 protein"
XX	PD	22-OCT-1998.	
XX	PF	16-APR-1998;	98WO-US07573.
XX	PR	16-APR-1997;	97US-084659.
XX	PA	(ARCH-) ARCH DEV CORP.	
XX	PI	Leopardi R, Rolzman B;	
XX	DR	WPI: 1998-594559/50.	
XX	DR	P-PSDB: AAM80810.	
XX	PT	Use of herpes simplex virus U(s)3 polypeptide - for developing	

PT products for modulating apoptosis in cells and for identifying  
 PT compounds which act as stimulators or inhibitors of apoptosis  
 XX  
 PS Example 2: Pages 60-63; 85pp; English.

CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in  
CC the method of the invention as modulators of apoptosis. The methods  
CC and products can be used to identify compounds which modulate  
CC (stimulate or inhibit) apoptosis in cells. They can be used to  
CC immortalise cells for the study of these cells or for growing cells  
CC in large numbers for the productions of proteins. They can also be  
CC used for stimulating apoptosis in cells, e.g. for treating a subject  
CC with a HSV infection.

SQ Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T; 0 other;

Query Match	8.5%	Score 84.6	DB 19	Length 4257
Best Local Similarity	42.9%	Pred. No. 2.2e+05		
Matches 423	Conservative 0	Mismatches 564	Indels 0	Gaps 0

[illegible]

[illegible]

Sequence	Location/Qualifiers
AAV10362/C	361..4257
AAV10362 standard; CDNA; 4257 BP.	/*tag= a
AAV10362;	/product= "Infected cell protein"
30-JUN-1998 (first entry)	
Infected cell protein number 4 alpha-4 gene.	
Infected cell protein number 4; ICP4; alpha-4; cell apoptosis; therapeutics; ss.	
Herpes simplex virus.	
Key	
CDS	
MO9804709-A2.	
05-FEB-1998.	
23-JUL-1997;	97MO-US12904.
26-JUL-1996;	96US-0690473.
(ARCH-) ARCH DEV CORP.	
Leopardi R, Roizman B;	
WPI; 1998-130697/12.	
P-PSDB; AAM40200.	
Use of herpes simplex virus ICP4 polypeptide - useful for, e.g. blocking apoptosis in cells, production of proteins and gene therapy	
Disclosure; Fig 2; 63pp; English.	
The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block cell apoptosis. Similarly the administration of an agent that inhibits ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells. This can be used for the immortalisation of cells, production of proteins, gene therapy, or inhibition of cell death induced in vivo. They can also be used for production of therapeutics comprising inhibitors of HSV ICP4 function, useful for treating HSV function.	
Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T; 0 other;	

Query Match	8.58;	Score	84.6;	DB	19;	Length	4257;
Best Local Similarity	42.9%;	Pred. No.	2.2e-05;				
Matches	423;	Conservative	0;	Mismatches	564;	Indels	0;
						Gaps	0;

QY	10	ggtctccggcgccgagccgagccgagccacacgccggggaagatcctgcctcagttcggaaag	69
Db	2490	ggcgcctccggctcgtccgcccccccccgcgcacaaagggggccacagctctggtctcmaag	2431
QY	70	cgcctgcctgcctccgaaggttcaatcgccggaacgcctctccagccgaggaagtgcctgcggaggt	129
Db	2430	caggtcccgccggccggccggccggccggagctccggcacaagccggcggtcccgccgacgcgggg	2371
QY	130	tcggcgagccgatgacgagcgccggccgagatgcggggccggggaacagaccatgagaccccaac	189
Db	2370	gcccaagggccccccggccgcacagagctcaacggggccacagggggccacggccgctctccctgc	2311
QY	190	ctcggccggggtctctccgacgaaggttcgcgcgaagaaattctctgatgactcgcgtccggcg	249
Db	2310	gccggccacccgcgcagctccccccggccgacagggccatgacacacacagccggtggccgcacaa	2251
QY	250	cagggaacgagcaaacagcaactcgcgcagcgccctgcgcgcgaattcaacggcggtgccggcg	309
Db	2250	cagctccgcgcacgcacacccacccgcacagggccggccgcgctgagctccggccggggagac	2191
QY	310	cagctctctccagcgctcgcctgccttgagacagcgccgcggttcagcgagcagcaagcaagcgc	369
Db	2190	ggggcccgccggagctccctccggccggccggggcgctggccggcccccggccacgccccgg	2121
QY	370	gaagcccgtagagcctgcgcgcgagcgagacacctccagctcgcgcgtgcgcgaacgaggtc	429
Db	2130	gaagcccccgcagagctccgctccctgcgaagccctccggcacaacgcctctcacagatccccgcacagcc	2071
QY	430	gtacagagcagctgcgtgcggagagacacatctccagagagatgctgcgcgagcgctctccctgcgc	489
Db	2070	ggccacagagactgcacagccgcacacagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	2011
QY	490	ggcgcgagatgctgcagagcccgccgaagtttcgacgagcccgagatgtagagacctgcgcgtgcgc	549
Db	2010	gtccgacgcggggggggccgctgctgggggctgctgcgcccccgcgggggagagccggccggccga	1951
QY	550	gaacagagcctgtccatgagtcgcgcgcacacacctgcctccacagcgcgagacggttcgatgtgtg	609
Db	1950	cagccgccccacagggccggccgcagagatcccccggcgccgctgacccgggggacccgcgcgctc	1891
QY	610	catctggttagagctgcgatgtgtgttcgcagagcccaagcctgcgcgagagatccctcgacagagagc	669
Db	1890	gcccggtccggccggccgcgcacagccggccgcaccccctgctgctgctgcgcgcgcgcgcgcgcgc	1831
QY	670	gatgatgagccgc	729
Db	1830	ccccggccggcccccctgcagggccgc	1771
QY	730	ctgtgttcgcagcaagcaggtgtgtctcgcgcgcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc	789
Db	1770	cagagctgctgctcagcaggaagcccttctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1711
QY	790	cagctcctcctggtgtgagcctctctagagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	849
Db	1710	ggccggccggccggccgc	1651
QY	850	gtctgatgcgcgcgtcccgagcgtgtgttcacatcaagcgagcgcgctgcgtcctcgcgcgcgcgc	909
Db	1650	ggggacacagcccccgcgcacagctccgc	1591
QY	910	tcgcgttaagttcaacgltgc	969
Db	1590	gattccggaagacagccctgcgttcacagccgcacagctccccggggacacagccgcgggtctgagag	1531
QY	970	ccgcagcgtgcgtcgcgagcgtcgtggt	996
Db	1530	ccaccccatggcctccgcctccgggct	1504

```

RESULT 10
AA53491
ID AA53491 standard; DNA; 114955 BP.
XX

```







16-JUN-1999; 99DE-1028313.  
16-JUN-1999; 99DE-1028313.  
(KELLER) KELLER U.  
Schaewecker F, Keller U;  
WPI: 2001-081744/10.  
Introducing N-methyltransferase activity into peptide synthetase  
enzymes, useful for synthesis of N-methylated peptides such as  
antibiotics, by altering DNA -  
Example 2; Page 9; 18pp; German.  
This invention describes a novel method of introducing an  
N-methyltransferase domain (A) into peptide synthetase (PPS) activation  
domains by altering the DNA that encodes the domains. The invention also  
describes a method for combining genes (or gene segments) that encode PPS  
modules that lack (A) with genes (or segments) encoding modules that  
include (A). Modified DNA sequences formed by insertion of (A) are used  
(1) for altering natural (or already altered) PPS or polypeptide  
synthetase (PKS) genes and their fragments, also for construction of new  
PPS and PKS gene and (11) for construction of plasmids or genetically  
altered organisms for synthesis of encoded proteins (B). (B) are used  
in vivo or in vitro enzymatic synthesis of amino acids, polypeptides, and  
peptidyl-acetyl mixed structures containing N-methylated amino acids, or  
these compounds, also for fermentative production of such compounds.  
These compounds are often of pharmaceutical value, e.g. penicillins,  
vancomycin, cephalosporin, pristinamycin or actinomycin D. The modified  
PPS (enzymes involved in non-ribosomal peptide synthesis) are able to  
N-methylate their substrates but substrate specificity remains unchanged

Query Match	8.1%;	Score 81.2;	DB 22;	Length 3849;
Best Local Similarity	44.0%;	Pred. No. 7.4e-05;		
Matches 399;	Conservative 0;	Mismatches 498;	Indels 9;	Gaps 1;

OY	63	cgagcagcgccctgtctgtccagaggtacatcgagacccctaaagccgagatcagctctc	122
Db	2274	CGGCTCGGACAGCGGGAGCGGCGGTATTCCTGGCCCTTTGGAGCGTGGTGTGTACGGCT	2215
OY	123	cgagggtctgcgaggccgaatgacgagccgagcgaatgcgcggagccgaggacagaccatgca	182
Db	2214	AGCGGGTCAACTCTGTTGGGCGGCGCCGACATTGAGCCGCAAGTGA-----CGC	2164
OY	183	gcccacactcgcgcggtctctgcgaggtttgcggaagactcttcgtabgacctgacg	242
Db	2163	CGGGAGGTCCGGGAGCGGGGTGGTGACCGCGGTGAAGTACCTCCGGTGTGACCAAGAAATT	2104
OY	243	ccggagcgagaggaagcaacagacacctgcgacagcgccctgcgcgaattcacgcggtgc	302
Db	2103	CTTTCACAGACACAGGCTTGTGCTCAGACGCGCGCGGACAGCGCGGCGGTGTGGCGGGGT	2044
OY	303	ccgcgcgcacgtctctcagcgctccgtctgaagcagggccgcgtgcagcgcgagcaagcgca	362
Db	2043	CTCTCGGCGCGGCGGTGTGGACGCGCGGTGTGAAGGTGTGGCAGACGCCCGGTGTTCGGA	1984
OY	363	agaagcccgagcccgtaggcctgcgcgagcgagagaccccaagctcgcggtgcgcgaagcg	422
Db	1983	TGTGCGCGATGACACAGCGCCCGCGGGGGCCACGACGCGCGCTGTCTGATGACTT	1924
OY	423	ccaggttgttcaagcacttggtaggaagcaatgcccagggagttggcgcgcgcgcttct	482
Db	1923	GGCGAGGATATGTGCGCTTTCGGGAAGTACGACACGAGTTGAGACACAGCGGTGTGGA	1864
OY	483	ccctgcgcgagcgatgtgtgcacagcccggaagtctgaagcagccgaagtctgaggaacctgc	542
Db	1863	ATGTGGCCACGCGGACCCCTCTGTGCTGTGGCGCGGCAACCGGCGCCGACAGGTGACCGCC	1804

[illegible]

RESULT	13
AAA58471/c	
ID	AAA58471 standard; DNA; 58857 BP

AC	AAA58471;
XX	
DT	31-OCT-2000 (first entry)
XX	

DE	Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30
XX	
KW	BIM gene cluster; bleomycin gene cluster; polyketide metabolites;
KW	bleomycin; bleomycin analogue; holo carrier protein; thiazolidine;
KW	thiazoline; dithiazoline; microbial metabolite; sugar; ss.
XX	
OS	Streptomyces verticillius.

	Location/Qualifiers
PH	Key
FM	CDS
FT	223..564
FT	/*tag= a
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 30; encodes AAB07556"
FT	561..2309
FT	/*tag= b
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 29; encodes AAB07557"
FT	2767..3486
FT	/*tag= c
FT	/note= "ORF 28; encodes AAB07558"
FT	3527..5593
FT	CDS
FT	/*tag= d
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 27; encodes AAB07559"
FT	5806..12294
FT	/*tag= e
FT	/note= "ORF 26; encodes AAB07560"
FT	12291..15491
FT	/*tag= f
FT	/note= "ORF 25; encodes AAB07561"
FT	15488..21013
FT	/*tag= g
FT	CDS

FT	/note= "ORF 24; encodes AAB07562"
FT	21010..24666
FT	/*tag= h
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 23; encodes AAB07563"
FT	24663..32690
FT	/*tag= i
FT	/note= "ORF 22; encodes AAB07564"
FT	32893..34830
FT	/*tag= j
FT	/note= "ORF 21; encodes AAB07565"
FT	34827..35804
FT	/*tag= k
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 20; encodes AAB07566"
FT	35818..37302
FT	/*tag= l
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 19; encodes AAB07567"
FT	37299..39215
FT	/*tag= m
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 18; encodes AAB07568"
FT	39301..47181
FT	/*tag= n
FT	/note= "ORF 17; encodes AAB07569"
FT	47178..49985
FT	/*tag= o
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 16; encodes AAB07570"
FT	49982..51001
FT	/*tag= p
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 15; encodes AAB07571"
FT	50998..52386
FT	/*tag= q
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 14; encodes AAB07572"
FT	52383..52946
FT	/*tag= r
FT	/note= "ORF 13; encodes AAB07573"
FT	53018..54190
FT	/*tag= s
FT	/note= "ORF 12; encodes AAB07574"
FT	54187..55824
FT	/*tag= t
FT	/note= "ORF 11; encodes AAB07575"
FT	55821..56093
FT	/*tag= u
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 10; encodes AAB07576"
FT	56090..57586
FT	/*tag= v
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 9; encodes AAB07577"
FT	57583..58857
FT	/*tag= w
FT	/note= "ORF 8; encodes AAB07578"
XX	
PN	W0200040704-A1.
XX	
PD	13-JUL-2000.
XX	
PF	06-JAN-2000; 2000MO-US00445.
XX	
PR	06-JAN-1999; 99US-0115435.
PR	05-FEB-1999; 99US-0118848.
PR	05-JAN-2000; 2000US-0447962.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
XX	Shen B, Du L, Sanchez C, Chen M, Edwards DJ;

[illegible]



```

QY 775 cagcacctgcacacagctctcgtgtgtgtgcccctgtatgacgcgca---gcgctacat 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23514 CAGGACCCCGGCGCTCGGGCGCGCCGGGAAGGTGATGAGGCCGCCACACCTCGGGCAG 23455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 832 gtccgagcgtgtgagcagctgtatgtatgcgcgcgcgtccgcgcgcgtgtcatcagcgcgcg 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23454 CTCACGCGGCGCACCCGGCCAGGCCCAACAGCTGGGCAACGGCGGTGCACGCGCAGTTTC 23395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 892 gtccgtctctcgcagcgcgtcgtgaattcaccgtgtccgcgcagacagcctgtgtgtgag 951
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23394 GTGCGGGGGGGGTGGGACCGCGCGCGGCACTGTGACGTGCCACAGCGGGGTGCTGCCGAG 23335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 952 cagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23334 GTGCGCGCAGGCGCTGTGACGACGAGCGGAGGTGGCGGTGACGCGCGC 23292

RESULT 15
AAT80414/C
ID AAT80414 standard; DNA; 44377 BP.
XX
AC AAT80414;
XX
DT 27-FEB-1998 (first entry)
XX
DE Platenolide synthase gene cluster.
XX
KM Tyalactone synthase gene cluster; tylg gene; multifunctional protein;
KW platenolide synthase gene cluster; platenolide production; stmG gene;
XX polyketide; tyalactone synthesis; antibiotic; tylosin; ss.
XX
OS Streptomyces ambofaciens.
XX
Key Location/Qualifiers
FH CDS 350..14002
FT /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT CDS /note= "ORF1 encodes protein shown in AAM22606"
FT 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAM22607"
FT 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT CDS /note= "ORF3 encodes protein shown in AAM22608"
FT 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAM22609"
FT 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAM22610"
XX
PN EP791655-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-0301056.
XX
PR 22-FEB-1996; 96US-0012078.
XX
PA (ELIL ) LILLY & CO ELI.
PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL.
XX
DR WPI: 1997-418046/39.
XX
DR P-PSDB; AAM22606-W22610.
XX
PT DNA encoding Streptomyces fradiae tyalactone synthase domain - for
XX production of tylosin-related polyketide compounds.
XX
Example 2; Pages 110-134; 220pp; English.

```

CC This sequence represents the platenolide synthase gene cluster of the  
CC isolated from Streptomyces ambofaciens. This sequence encodes the  
CC multi-functional proteins which direct the synthesis of the polyketide  
CC platenolide. Platenolide is the basic building block of the antibiotic  
CC antibiotic spiramycin. This sequence was used along with the tylg gene  
CC (see AAT80413) to create a hybrid ORI sequence (see AAT80415). The tylg  
CC gene is the tyalactone synthase gene cluster of the invention. The tylg  
CC sequence was isolated from Streptomyces fradiae, and encodes  
CC multifunctional proteins which direct the synthesis of the polyketide  
CC tyalactone. Tyalactone is the basic building block of the antibiotic  
CC tylosin. The tyld sequence can be used to transform S. ambofaciens  
CC lacking the stmG ORI sequence, or S. fradiae lacking the tylg ORI  
CC sequence, so that they can produce polyketides. The DNA sequence can be  
CC modified so as to alter the type of carboxylic acids incorporated, the  
CC number of carboxylic acids incorporated and/or the post-condensation  
CC reactions performed, thereby resulting in novel tylosin-related  
CC polyketides.

SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

Query Match 7.8%; Score 77.8; DB 18; Length 44377;  
Best Local Similarity 44.7%; Pred. No. 0.00018;  
Matches 448; Conservative 0; Mismatches 537; Indels 18; Gaps 3;

```

QY 7 ccaggtctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24291 CCACCTGACGTCGCGGACCGGACCGGACCGGCGCTGCGGCTCCAGATGCGGCGCAGGAT 24232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 cagcgcctcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24231 CCGCACCGCGCTGTCGCGGTCGAGGAGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 ggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24171 GGCACGCTGCGGCGGCGGCTGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 187 cactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24111 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 235 ctgcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24051 GCCCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 295 cgcgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23991 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23932
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 ccagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23931 GCGGCGGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23871 GTCCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 475 ggcgtctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23811 GACGTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 535 gacctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23751 GGGGGGCGTGGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 595 ccggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23691 CAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 655 tcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23631 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```







Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ctgcaacagcgtctccgagcgccgagcgagccagcgaggaatctcgtcag 60
Db 3407 CTGCAACAGGCTTCGGGGCCGGGGGAGGACCGGGGGAGATCTGTCAG 3348
QY 61 ttccgacagcgtctcgtccagggatcgcgagacgtctcagcgagtcagctg 120
Db 3347 TTCGAGAGGCGCTGCTGTCAGAGGTCATCGGAGCGCTTCAGCGGAGTCAGCTG 3288
QY 121 ctgaggggttcgagggcgatgacgagcgagcgatgacgagcgagcgagcgagc 180
Db 3287 CTCGGGGGTTGGGGGCGATGAGGGGCGGAGTCCGGGCGGAGCACACCTATGC 3228
QY 181 gaagccaccctcgccgaggtcttcgagaggttgcgagcagaacttcctagagcctgat 240
Db 3227 GAGCCCACTCGGCGGGCTTCGCGGAGTTGGCGCAACACTTCGTAAGGCTTGAT 3168
QY 241 cgcgcgagcgagggagcgagcaacagcactgcgaagcgccctgcgcgaacttcacgcggt 300
Db 3167 CGCGGGGCGGAGGAGGACGGAACAGCACTCGCACGCGCTCGCGCCGACTTCACCGCGGT 3108
QY 301 gccgcgagcgaggtcttcgagcgctcgcgtgcgagcgagcgagcgagcgagcgagc 360
Db 3107 GCGCGGCGGAGGCTTCGAGCGCTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3048
QY 361 gaagacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 420
Db 3047 GAAGACGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 2988
QY 421 cgcgcgaggttgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 480
Db 2987 CGCGAGGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2928
QY 481 ctccgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 540
Db 2927 CTCCTCGGCGGCGGCGGAGTGCAGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2868
QY 541 gccgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 600
Db 2867 GCGCGTGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2808
QY 601 gtagtgatctgtagcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 660
Db 2807 GATGTGTCATCTGTAAGAGTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 2748
QY 661 gtaggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 720
Db 2747 GCAAGAGGCGATGATGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2688
QY 721 gccgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 780
Db 2687 GCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2628
QY 781 ctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 840
Db 2627 CCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2568
QY 841 gtaggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 900
Db 2567 GTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2508
QY 901 ctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 960
Db 2507 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2448
QY 961 acgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 999
Db 2447 ACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2409
```

RESULT 2  
US-08-690-473-1/c

```
; Sequence 1, Application US/08690473
; Patent No. 5876923
; GENERAL INFORMATION:
; APPLICANT: Leopardi, Rosario
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICP4 AS AN
; TITLE OF INVENTION: INHIBITOR OF APOPTOSIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,473
; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4257 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-690-473-1
```

Query Match 8.5%; Score 84.6; DB 2; Length 4257;

Best Local Similarity 42.9%; Pred. No. 1.3e-07;

Matches 423; Conservative 0; Mismatches 564; Indels 0; Gaps 0;

```
QY 10 ggtcttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 69
Db 2490 GCGGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2431
QY 70 cgcctgctgctcagggatcagggagcgagcgagcgagcgagcgagcgagcgagcgagc 129
Db 2430 CAGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2371
QY 130 tcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 189
Db 2370 GCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2311
QY 190 ctcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 249
Db 2210 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2251
QY 250 cagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 309
Db 2250 CAGGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2191
QY 310 cagcttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 369
Db 2190 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2131
QY 370 gtagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 429
Db 2130 GAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2071
QY 430 gtacagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 489
```











OY	250	caggagacgacaacgaacactctgcgacggccctgcgcgcgaattcacacgcggtgcgccgcgcgc	309
Db	4424	CAGCTCCGCGACACCACCGCGCACGGCGGGGCCGCTGCCTCGCGCGCGCGGGGAAGC	4483
OY	310	cagcttctccagcgcgtctcgcttgacgcagcgccgcgttcgcagcgcgacgcagaagcgaaacgcgc	369
Db	4484	GGGGCCCCGGGTCCCTCCGCGCGCGGGGGGCTGGGGGGCCGCGCCCGGACAGCCCCGG	4543
OY	370	gagcccgtaggcctctgcgcgcgcgcgcgcgcacacctccagcttcgcgcgtgcggaaccgcagttc	429
Db	4544	GACGGCGCGCAGGTCGCCCTCGAAGCCCTCGCCGCGCAAGCGCTTCAAGATCCCGCGGACAGGC	4603
OY	430	gtaacgcactctgttgaggagacatcgtcccaagagagtgcgcgcgcgcgcgcgcgtctctctgcgc	489
Db	4604	GCCCAAGGCAATTCAACAGGCCAACGCGGCCGCTGTGGCGCGCGGCCCTCCGCGGTGTTGGGC	4663
OY	490	ggcgcgcgatgtgcacagcccgc	549
Db	4664	GTCGCGCTGGCGGGCGCGCGCTGCGGGGTCTGTCGCCGCCCGCGGGGAGGCGGCGGGCGGA	4723
OY	550	gagcagagctgtccatgctgcctgtgcacacactcgtcccaagcgcgcgacgcgcgcgcgcgcgcgc	609
Db	4724	CA-GCCCCCACAAGCGCGCGAGAATCCCCTCGCGCGCTGTAACCGCGGCGCACCGCGCGCTC	4782
OY	610	catctgttagacgtcgtgatgtgctgc	669
Db	4783	GCCCCGATGCGCGCGCGCGCGCACGGCGCGCAACCCCTTGCTATATTGCGCGCGCGCGGGGCT	4842
OY	670	gatgatgtgc	729
Db	4843	CCCCGCGCGCCCCGTGACGCGCCCGCTTTCTTCGCGCGCAACAAGGCGCGCGGTATGGCGCGCG	4902
OY	730	cttgtgtgc	789
Db	4903	CAGGCTGTGATAGAGAAAGCCCTTTGCGCGCGCGGTGATTCGCGCGCTCATGCGCACAGGC	4962
OY	790	cagctctcctgcgtgtgcgcctcttgagagc	849
Db	4963	GCGCGCGCGCTGCGCGCGACGAGGCCCAAGCGGAAGGGCGCGCGCGCGCATAGCGGTATGCCAGGTG	5022
OY	850	gttatgtgc	909
Db	5023	GCGCACGCGCGCGCGCACGCTGCGCGGTATGAAGAGTGCTGTTCGCGCGCGCGCGCGA	5082
OY	910	tccgctgaatatccatccgttcgc	969
Db	5083	GATCCGAGAACAGGCGCTGTGTCAGGCCCAAGTCCCGCGGAGACACGCGCGGTTCTGGAG	5142
OY	970	ccgcgcgtgcgtgc	996
Db	5143	CCACCCCATGGCTTCGCGCTCGCGGGCT-5169	

US-08-804 8  
US-08-804-227C-1/C  
Sequence 1, Application US/08804227C  
Patent No. 5876591  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostreck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM Compatible
3 OPERATING SYSTEM: MS-DOS
4 SOFTWARE: ASCII(DOS) text only
5
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/804.227C
8 FILING DATE: February 21, 1997
9
10 CLASSIFICATION: 435
11
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Plant, Thomas, G.
14 REGISTRATION NUMBER: 35,784
15 REFERENCE/DOCKET NUMBER: X-8231
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: 317-276-2459
18
19 INFORMATION FOR SEQ ID NO: 1:
20
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 43280 base pairs
23 TYPE: nucleic acid
24 STRANDEDNESS: single
25 TOPOLOGY: linear
26
27 MOLECULE TYPE: DNA (genomic)
28
29 FEATURE:
30 NAME/KEY: CDS
31 LOCATION: 816..14234
32 FEATURE:
33 NAME/KEY: CDS
34 LOCATION: 14351..19945
35 FEATURE:
36 NAME/KEY: CDS
37 LOCATION: 20010..31199
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: 31232..36067
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: 36249..41774
44
45 US-08-804-227C-1

```

[illegible]





OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Infringement #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/510,646B  
 FILING DATE: 03-AUG-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/403,852  
 FILING DATE: 10-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/FR 93/00923  
 FILING DATE: 25-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-01000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5392 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: S.pristinaespiralis  
 US-08-510-646B-1

Query Match	7.1%	Score 71.2	DB 3	Length 5392
Best Local Similarity	45.4%	Pred. No.	3.2e-05	
Matches 305	Conservative 0	Mismatches 358	Indels 9	Gaps 1

```

OY      887 agcgatcgctgc 898
      1 1 1 1 1 1
DB      2687 ATTCCTTCGTG 2676

RESULT 12
US-08-510-646B-1/c
; Sequence 1, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanché, Francis
APPLICANT: Crouzel, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patrick
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Creely-Iagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pinegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

[illegible]





